Modelling Bacterial Growth for Applying Photodynamic Therapy with Indocyanine Green

Minh Nguyen

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 05.03.2018

Thesis supervisor:

Docent Kai Zenger

Thesis instructor:

Prof. Jarmo Alander

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Minh Nguyen

Title: Modelling Bacterial Growth for Applying Photodynamic Therapy with Indocyanine Green

Date: 05.03.2018 Language: English Number of pages: 10+56

Department of Electrical Engineering and Automation

Professorship: Control Engineering

Supervisor: Docent Kai Zenger

Instructor: Prof. Jarmo Alander

Antimicrobial approaches using photodynamic therapy (PDT) have become popular in medication. However, to the best of our knowledge, no model has been developed for estimating the dose of light and photosensitiser with respect to bacterial inhibition.
This thesis aims to model the growth of *Escherichia coli* which can be utilised in developing the aforementioned model when using PDT with near-infrared (NIR) light and indocyanine green (ICG).
The project applied a spectroscopic method to measure the spectra of bacteria and chemometric methods to analyse the spectral data. The project consists of two main phases. The first phase conducted two phantoms to develop a measurement system and identify the possibility of utilising the system in controlled conditions. The first phantom analysed the spectra of LED lights. The second phantom determined the concentrations of different colour liquids. In the second phase, bacterial suspension was subjected to spectral analysis using the developed system. As a result of the whole project, a model has been established in order to monitor and estimate the concentrations of bacteria in liquid samples.

Keywords: modelling, bacteria, spectrocopy, principal component analysis (PCA), partial least squares regression (PLSR)

Preface

Firstly, I would like to greatly thank Docent Kai Zenger for giving me the opportunity to work on this Thesis. He has been a very patient supervisor who cared about my thinking and has helped ease my stress when tackling with difficulties during the project. He has always supported me to understand the problem deeply and given me chances to practice expressing my opinions.

Secondly, I want to thank Professor Jarmo Alander for his excellent guidance and support during this process. He has instructed me throughout this thesis and introduced me to the application of automation in biology. He has also continuously supported and advised me possible solutions to tackle the problems. Also, thanks for his stories about the industrial world and his experience.

I want to say big thanks to Marika Hellman as my instructor toward the tiny microbiological world. She has guided me through the basic biological lab skills as well as never lost her patience with the questions of an amateur like me. She has also shown me plenty of interesting aspects of biology which I have never acknowledged before.

I also would like to thank my friends who have assisted me to understand the theory of this work, who have supported in improving my paper, and who have always motivated me. I also benefited from debating issues with them to comprehend the problems.

Finally, I owe my parents a debt of gratitude. They have always encouraged me as well as given me wise counsel. You deserve my particular thanks and loves.

Otaniemi, 05.03.2018

Minh Nguyen

Contents

List of Figures

List of Tables

Symbols and abbreviations

Symbols

i	dummy index for a sample or an observation
j	dummy index for a vector
h	dummy index for components
k	dummy index
A	absorbance
I	intensity of light
ϵ	absorptivity of the sample
c	concentration of the sample
l	optical path length
f_{vib}	bond frequency of absorption band
k_b	force constant of the bond
μ	reduced mass of the bonded atoms
E_{vib}	vibrational energy
h_P	Planck's constant
v	vibrational quantum number
H	radiance exposure
n	refractive index of transmitting medium
ν	size parameter
r_p	dimension of the object, for spherical object r_p is the radius
λ	wavelength of light
t	exposure time
$\mathbf{X}$	data matrix or a matrix of features for independent variables ($\mathbf{X} \in \mathbb{R}^{N \times p}$)
$\mathbf{X}'$	transpose of matrix $\mathbf{X}$
$\mathbf{x}^{(i)}$	row vector in $\mathbf{X}$ from data of an observation
$\mathbf{x}_j$	column vector of an independent variable in $\mathbf{X}$
$\mathrm{x}_j^{(i)}$	an element of $\mathbf{X}$
η_j	mean value of variable x_j
$\mathbf{x}_i'$	transpose of vector $\mathbf{x}_j$
$\mathbf{H}$	transformed data matrix ($\mathbf{H} \in \mathbb{R}^{N \times p}$)
$\mathbf{Y}$	matrix of dependent variables $\mathbf{H} \in \mathbb{R}^{N \times q}$
N	the number of observations or samples
m	the number of principal components of $\mathbf{X}$
p	the number of independent variables or features
q	the number of dependent variables
r	the number of principal components of $\mathbf{Y}$
$\mathbf{T}$	matrix of scores of $\mathbf{X}$ ($\mathbf{T} \in \mathbb{R}^{N \times m}$)
$\mathbf{P}$	matrix of loadings $\mathbf{X}$ ($\mathbf{P} \in \mathbb{R}^{p \times m}$)
$\mathbf{P_t}$	matrix of full loadings $\mathbf{X}$ ($\mathbf{P_t} \in \mathbb{R}^{p \times p}$)

$\mathbf{t}_h$	column vector of scores for the h-th component
$\mathbf{p}_h$	column vector of loadings for the h-th component
$\mathbf{S_X}$	covariance matrix of $\mathbf{X}$
$\mathbf{S_H}$	covariance matrix of $\mathbf{H}$
$\mathbf{\Gamma}$	a diagonal matrix including eigenvalues of $\mathbf{X'X}$
$\mathbf{D}$	unitary matrix resulting from singular value decomposition
$\mathbf{\Sigma}$	diagonal matrix of singular values
σ^2	variance of a data set
$\mathbf{V}$	unitary matrix resulting from singular value decomposition
$\mathbf{d}_j$	column vector of $\mathbf{D}$
$\mathbf{v}_j$	column vector of $\mathbf{V}$
$\mathbf{B}$	coefficient matrix
b_h	regression coefficient in $\mathbf{B}$
$\mathbf{U}$	matrix of scores of $\mathbf{Y}$ ($\mathbf{U} \in \mathbb{R}^{N \times r}$)
$\mathbf{Q}$	matrix of loadings of $\mathbf{Y}$ ($\mathbf{Q} \in \mathbb{R}^{q \times r}$)
$\mathbf{E}$	matrix of noise term for $\mathbf{X}$ ($\mathbf{E} \in \mathbb{R}^{N \times p}$)
$\mathbf{F}$	matrix of noise term for $\mathbf{Y}$($\mathbf{E} \in \mathbb{R}^{N \times q}$)
$y^{(i)}$	true concentration of a sample
$\hat{y}^{(i)}$	predicted concentration of a sample
R^2_C	the coefficient of determination of calibration
R^2_{CV}	the coefficient of determination of cross validation
$\mathbf{B}_w$	weighted regression coefficient matrix
β	coefficient in $\mathbf{B}_w$
$\mathbf{W}$	the weight loadings of $\mathbf{X}$ ($\mathbf{W} \in \mathbb{R}^{p \times m}$)
k	the number of data subsets
s_j	standard deviation of variable x_j
rpm	revolutions per minute

Abbreviations

ALS	Alternating least square
ANN	Artificial neural network
AU	Absorbance units
CFU	Colony forming units
E. coli	*Escherichia coli*
GA	Genetic algorithm
GRAM	Generalized rank annihilation method
ICG	Indocyanine green
ILS	Intermediate least square regression
IVS	Iterative variable selection
LDA	Linear discriminate analysis
LV	Latent variable
MLR	Multiple linear regression
NIPALS	Nonlinear iterative partial least squares
NIR	Near-infrared
NIRS	Near-infrared spectroscopy
OD	Optical density
OPA	Orthogonal projection analysis
PLS	Partial least squares
PLSR	Partial least squares regression
PC	Principal component
PCA	Principal component analysis
PDT	Photodynamic therapy
PRESS	Predicted residual error sum of squares
PS	Photosensitiser
RMSEC	Root mean square error in calibration
RMSEP	Root mean square error in prediction
RMSECV	Root mean square error in cross validation
SIMCA	Soft independent modelling of class analogy
SNR	Signal-to-noise ratio
RAFA	Rank annihilation factor analysis
ROS	Reactive oxygen species
rpm	Revolutions per minute
SMA	SubMiniature version A
SVD	Singular value decomposition
TVC	Total viable count
VIS	Visible
UV	Ultraviolet
WRC	Weighted regression coefficients

1 Introduction

Since the late 1980s, a global crisis of antibiotics resistance has radically grown due to two significant phenomena: the decreased introduction of new antibiotics as well as the increased wide misuse and overuse of medication have increased widely. This has resulted in pathogens gradually developing an immunity against previously effective antibiotic drugs, and in turn steadily causing untreatable illnesses [1][2]. Immunisation is exacerbated when more drugs are used because it causes microbes to replicate and to create new mutants which resist antibiotics [3]. Hence, an urgency has arisen to develop new aseptic approaches, especially ones without drug usage.

One of these approaches, photodynamic therapy (PDT) has been used in numerous medical treatments to fight infectious diseases [4]. PDT utilises a light with a specific wavelength to excite a photosensitiser (PS), a chemical molecule which absorbs light and is also known as a dye. The excited molecule alters chemically into another molecule. This therapy elicits cell death through a sequence of photochemical and photobiological processes [5]. Hence, it can be used in antimicrobial treatments, killing microorganisms or stopping their growth. One solution to this therapy is a combination of near-infrared (NIR) light and indocyanine green (ICG). ICG is a low toxic PS and can be rapidly excreted from the human body. It was authorised by the United States Food and Drug Administration in 1956 and has been commonly employed in medical diagnoses, such as retinal angiography, liver clearance test, and the observation of blood vessels [4][6]. Recently, ICG has been widely applied in intra-operational imaging in surgery, cancer treatment, and a broad variety of other clinical applications [6]. One of its properties is absorbing near-infrared light. The absorption peak is reached at a wavelength of approximately 805–810 nm [7]. When absorbing NIR light, ICG can excite oxygen from the ambient environment and lead to the production of singlet oxygen [4][8]. Singlet oxygen is capable of inhibiting the viability of bacteria, thus functioning as an anti-infectious agent [9].

Despite its vast applications, health treatment based on the PDT applying ICG, particularly antimicrobial applicability, still needs further development to overcome some undesired effects. On the one hand, PDT typically uses a laser as the main light source. The laser energy usually possesses a risk of thermal damage. On the other hand, when the concentration of the PS or the power of the light is low, the effect of killing bacteria or tumour can be insufficient, thus activating bacterial proliferation, or requiring multiple treatments [4]. Hence, identifying an appropriate dose of ICG and power of NIR light is meaningful to medical applications.

A great deal of previous research has concentrated on investigating the effect of PDT applying ICG on anti microbes and antitumour. However, none of them has associated the concentration of indocyanine green (ICG) and the power of near-infrared (NIR) light with the number of viable bacteria. This thesis aims to develop a model for the bacterial growth which can be used in establishing the aforementioned association. This work utilises a spectroscopic method since it can both quantify bacteria and measure the power of NIR light in the future development. The model of this work will correlate the spectral response and the concentration of bacterial cells in the suspension.

This thesis work was implemented as an empirical project with two phases. In the beginning, two phantom experiments were conducted using LEDs and colour liquids as instances to implement spectral analysis and to estimate concentration, respectively. In the second phase, the number of bacteria is quantified utilising the developed spectroscopic approach from the first step. The quantitative results are extracted by applying chemometric tools as they are broadly used in analysing chemical and biological information [10]. One of the most extensively used tools is partial least square (PLS) regression which correlates the concentrations of each constituent with their spectral responses. It has been widely utilised not only in chemical analyses but also in bacterial discrimination and quantification [11][12]. Spectroscopic methods only acquire spectra; hence, they require true concentrations from a reference method to establish a model. In this project, the optical density and the traditional plating method are selected due to their popularity in microbiology.

This thesis has been organised as follows. Chapter 2 describes biological and medical theories in the mechanisms of PDT to induce singlet oxygen and its effect on damaging cells. Chapter 3 illustrates the mechanisms of spectroscopy based on light attenuation due to absorption and scattering as a method for quantifying bacteria. Chapter 4 provides two methods of chemometrics implemented in this project including principal component analysis (PCA) and partial least square (PLS) for data analysis. Chapter 5 demonstrates the two phantom experiments implemented for setting up the measurement system, acquiring spectra, and applying basic data analysis tools. Chapter 6 presents the empirical procedure with bacteria including sample preparation, reference methods, spectral acquisition, and data analysis. Chapter 7 demonstrates the results of the bacterial suspension measurement with important discussions. Finally, conclusions are presented in Chapter 8.

2 Biological and medical background

This chapter summarises the biological and medical background theories which motivated this project and will be used in future development. The first core premise is photodynamic therapy, which utilises a photosensitiser and light to inhibit bacterial growth, particularly indocyanine green (ICG) and near-infrared (NIR) light in this study. Their physical and chemical properties are described in accordance with their influence on the project. Secondly, reactive oxygen species (ROS), particularly singlet oxygen, which have been proved to be toxic to microbes, are presented to briefly clarify the reason for using PDT. Finally, the undesired effects of PDT are illustrated as the reasons behind conducting this research.

2.1 Photodynamic therapy

In the antimicrobial sector, photodynamic therapy employs a non-toxic photosensitiser (PS), such as toluidine blue, methylene blue, or ICG [8], and light at visible or near-infrared spectrum to elicit oxygen radicals leading to cell death. During this procedure, PS absorbs light to excite oxygen-containing agents and creates ROS which are deleterious to microbes as described in Section 2.2. PDT functions follow two types of mechanisms:

- Type 1: PS is excited to interact directly with the organic substrates of the cells. This process generates radicals (atoms or molecules include unpaired valence electron) and radical ions such as hydroxyl radical and superoxide anions. Those products react with oxygen molecule and create cytotoxic species such as hydrogen peroxide and superoxide which can destroy the tissues or cells. [4][13]
- Type 2: PS absorbs energy and transfers it to molecular oxygen which presents in the surrounding environment. This process provokes the formation of reactive intermediates of oxygen including singlet oxygen which is also detrimental to cells as demonstrated in Section 2.2 [4][13]. This mechanism is adopted in this study using ICG as the PS.

In fact, this therapy has been continuously studied and applied to multiple types of treatments. For instance, it treats pathogen infections which can cause serious health problems [4][9], prevents biofilm formation in dental diseases [8][14], cures dermatology such as acne [15] or melanoma [16], treats various cancer types such as colonic cancer [17] or breast cancer [18], and numerous others.

2.2 Reactive oxygen species and singlet oxygen

Reactive oxygen species (ROS) are highly reactive chemical molecules originating from molecular oxygen. Basically, ROS are formed through the addition of electrons to the oxygen molecule which consists of two unpaired electrons in the outer electron shell. This process produces various oxidising agents (radicals) such as singlet oxygen, superoxide, hydrogen peroxide, hydroxyl radical, hydroxyl ion, nitric oxide, and ozone [19][20]. Since they contain unpaired electrons, they tend to capture electrons from other molecules through chemical reactions. Some of those products are cytotoxic to biological creatures, meaning they are toxic to living cells. The reason is that they prevent natural detoxification mechanism. Detoxification is the process of removing poisons from the cells. When the level of radicals exceeds their detoxifying ability, cells are prone to death through activation of the apoptosis pathway [9]. Additionally, the radicals break indispensable macromolecules such as proteins, nucleic acids, and lipids, which play essential roles in constructing the cells. Despite being noxious, these properties of ROS can be beneficially derived in antibacterial and antitumour applications, thus they have been extensively studied in research and utilised in the medical field [9][13].

Singlet Oxygen, denoted as $^1O_2$, is a type of highly reactive ROS. It is the first excited state of an oxygen molecule which possesses a higher level of energy compared to a triplet oxygen [9][13]. Hence, it is unstable and easily oxidises other molecules. Singlet oxygen is acutely lethal to microorganisms in a similar manner to other ROS [9][13]. However, diverse types of bacteria as well as the wall structures of the cells have different susceptibility to this agent. For example, based on the structures of the cells, microbiology divides bacteria into gram-negative and gram-positive groups. Gram-negative bacteria have lipopolysaccharide membrane, thus their initial protection for the cell from extracellular singlet oxygen is higher than gram-positive ones [21]. This outer membrane prevents the penetration of singlet oxygen into cells as the molecules must pass the coat before attacking the vital components of the cells such as cytoplasm [21]. Furthermore, the membrane also plays the role of trapping singlet oxygen because it is made from organic molecules which can easily form chemical reactions with singlet oxygen [21].

2.3 Limitations of PDT with NIR and ICG

Photodynamic therapy utilising near-infrared light (NIR) and indocyanine green has been proved to induce almost no side effects on the host cells. However, NIR light can cause damage to healthy tissues, injure retina if the irradiation overdoses. NIR is an invisible region of electromagnetic radiation covers the range from 760 nm to 1400 nm (close to the visible region) as can be seen from Figure 1 [10]. The light is

actively absorbed by water molecules. This radiation can penetrate to the deep layers of tissues and be absorbed by the water molecules in skin tissues [22]. This increases the temperature of water and heats the irradiated tissues. Additionally, in PDT, NIR light is usually produced from diode lasers which possess a high energy level and can accelerate the heating process. During PDT, the temperature around the treated area can rise approximately 8 °C. If the ambient temperature reaches 43 °C, tissue can be irreversibly damaged [23]. This thermal injury, however, depends significantly on the exposed area, perfusion, pigmentation and initial skin temperature. The following equation can be applied to calculate radiance exposure H determined for wavelengths less than 3000 nm [22]:

$$H = 2.0 \cdot t^{0.25} \cdot 10^4 \tag{1}$$

where t is the exposure time. The length of time t should be limited less than 10 seconds to prevent thermal injury [22]. This undesired effect should be carefully considered when implementing PDT. On the other hand, if the power of light is insufficient, the killing effect of PDT may be inadequate which requires multiple treatments [4].

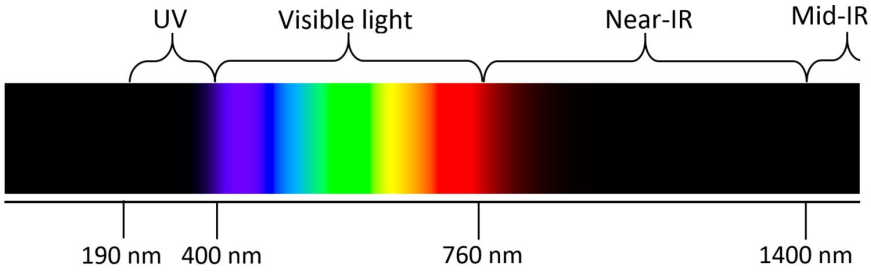


Figure 1: Light spectrum [10].

3 Spectroscopy for bacteria monitoring

Spectroscopy is a typical approach used in radiance measurement as well as chemical and biological analyses. Spectroscopy is applied in this project because of its capability to quantify bacterial cells and the feasibility to measure light power in a future study. This chapter will describe its working principles based on the attenuation of light when traversing through a medium. This phenomenon can be explained by the absorption and scattering of the incident light caused by the particles existing in the medium [24][25]. This manner is applied in spectroscopy for estimating concentrations. In numerous cases, the two effects cannot be clearly separated from each other [25]. This chapter includes three sections. The first two sections present the theories of those mechanisms. The last section illustrates their applications in spectroscopic approaches used for quantifying bacteria.

3.1 Absorption spectroscopy

UV-VIS-NIR spectroscopy (ultraviolet–visible–near infrared) is based on the fundamentals of light absorption. Absorption spectroscopy adopts the theory of energy transfer between light and matters. Light is separated into distinctive regions (spectrum) featured by different wavelengths. Photons of each wavelength have a specific frequency as well as a corresponding energy level. Thus, light photons can trigger energy transmission to molecules. In normal condition, molecules stay in a stable energy level. They are excited to a higher energy level when receiving an amount of energy which equals to the difference between any two of their energy states. When a photon collides with another matter, it transfers its energy to the matter. If the amount of energy of the photon is relevant to a particular energy transition of the atom (the difference between two energy levels), the atom is excited to a higher state. The process is known as absorption. Depending on the type of atoms that received energy, distinguished spectra will be emitted from the system after the reaction. Equivalently, absorption band properties can provide unique information on the physical and chemical characteristics of different substances and materials.[10]

Absorption spectroscopy can be disclosed through harmonic oscillation and resonance theory [10]. In classical physics, molecular vibration can be described as the model of a harmonic diatomic oscillator. In this system, two atoms connected through a bond with bond constant k_b can be considered as two masses connected through a spring with a similar constant. According to Hook's law and Newton's law, the oscillation frequency can be calculated as:

$$f_{vib} = \frac{1}{2\pi}\sqrt{\frac{k_b}{\mu}} \tag{2}$$

in which f_{vib} presents the bond frequency of absorption band, k_b signifies the force constant of the bond, and μ denotes the reduced mass of the bonded atoms. Vibrational energy can be deduced with Schrodinger equation:

$$E_{vib} = h_P f_{vib}\left(\mathrm{v} + \frac{1}{2}\right) = \frac{h_P}{2\pi}\sqrt{\frac{k_b}{\mu}}\left(\mathrm{v} + \frac{1}{2}\right) \tag{3}$$

where v is the quantum number (0, 1, 2, ...) of the vibration, h_P is Planck's constant ($6.6261 \times 10^{-34}\mathrm{m^2kg/s}$). From this equation, vibrational frequency and energy are functions of bond strength characterised by the structure of the compound [10].

UV-VIS absorbance spectroscopy is a rapid and reliable analytical tool applied in detecting, identifying and quantifying chemical substances as well as microorganisms and cells. UV radiation spans from 190–400 nm and visible light covers the range from 400–760 nm [10]. The absorption of a photon in these ranges excites an electron to a higher energy orbital. Then, the electron returns to its ground state v = 0 through fluorescence. The quantification of absorption data in these regions is frequently analysed with Beer-Lambert law. It states that the attenuation of radiant power of the light beam is proportional to the concentration of the absorbing species and to the length of the optical pathway [25]. The law is illustrated with the formula [25]:

$$A = -\log_{10}\frac{I}{I_0} = \epsilon(\lambda) * c * l \tag{4}$$

where A is the absorbance [AU], I is the intensity of light received by the photo detector after transmitting through the sample [counts], I_0 is the incident intensity of light [counts], $\epsilon(\lambda)$ is the absorptivity or attenuation coefficient of the species at the wavelength λ, c is the concentration of the sample, and l is the optical path length [m]. Generally, most spectrometers are standardised with a path length of 1 cm using a relevant cuvette and the absorbance is then signified with optical density (OD) [25].

UV-VIS spectroscopy is beneficial for analysis since the spectra consist information of the samples such as number, size, shape, and chemical composition of the suspended particles [26]. UV-VIS spectroscopy is less likely to cause heating risk to the measured targets. On the other hand, UV-VIS spectroscopy technique possesses some disadvantages. Traditionally, UV-VIS spectroscopy requires performance with diluted samples, or a short optical path length. Thus, it needs sample preparation.[10] This technique can be easily affected by scattering [27]. The scattering information is always contained in the spectrum and no mathematical means is capable of eliminating or distinguishing it. It is also disadvantageous due to dependence on the ambient factors such as temperature, pH, and viscosity [27]. Additionally, UV-spectroscopy possesses a high photon energy level and can be adverse to living organisms.

Near-infrared spectroscopy (NIRS) is an easy-to-implement, no-sample-preparation and non-invasive-interact technique that has been utilised for physical characterisation and chemical constituent analyses in a variety of compounds [28]. It has been widely applied in a diversity of fields such as pharmaceutical industry, raw material testing, remote sensing, food quality controlling, process monitoring, industrial and medical imaging, forensics, crime detection, and military [28]. NIRS extends over the electromagnetic spectrum from 760 nm to 2500 nm. As a part of light spectrum, NIR light stimulates the vibration frequency of the molecules. The technique analyses the information from absorption spectra of NIR based on two following modes [10]:

- ***Overtone*** From equation (3), it can be concluded that energy levels distribute discretely. Thus, overtone transitions are ones with the quantum number v greater than one (i.e. multiples of the fundamental frequency).
- ***Combination vibration*** This type of vibration happens in polyatomic molecules including different types of molecules. Each molecule needs to receive a different level of energy and then vibrates at different frequencies. Hence, this mode is the sum of multiples of each interacting frequency.

NIRS includes the first overtone, the second overtone, the third overtone and the combination band regions as shown in Table 1. The most significant absorption bands in NIR region relate to the fundamental molecular vibrations of some bonds such as C–H, N–H, O–H, and S–H. Therein, they can uniquely specify most chemical and biochemical species [10]. Consequently, they can be used for qualitative and quantitative analysis.

Table 1: Wavelengths in nm of some absorption bands of organic compounds [10].

Wavelength (nm)	**Assignment**
450–550	combination S–S stretching
600–700	combination C–S stretching
775–850	third overtone N–H stretching
850–950	third overtone C–H stretching
950–1100	second overtone N–H stretching; second overtone O–H stretching
1020–1060	combination S=O stretching
1100–1225	second overtone C–H stretching
1300–1420	combination C–H stretching
1400–1600	first overtone N–H stretching; fist overtone O–H stretching

NIRS is advantageous in the industrial and academic research environment. Typically, its spectrum provides more information than the normal electromagnetic spectra. Its absorption band is 10–100 times weaker than of the mid-IR region, hence it supports

quick inquiries into strongly absorbing or highly light-scattering matrices such as suspensions, pastes, and powders. It requires no reagents, no additional chemicals, and can be used in water-free applications. Due to the deep penetration ability of NIR, the samples need no dilution (i.e. reducing the concentration of the liquid) [28].

However, some disadvantages of NIRS need to be considered carefully while implementing experiments. Firstly, NIR spectra are normally complex and suffer from broad band overlapping, which results in a complicated process to treat the data. Secondly, every physical and chemical species present in the sample can affect the spectra. When the sample contains water, temperature is also a considerable factor which can shift the peak position influencing the accuracy of the measurement. Furthermore, it needs a reference method for calibration due to the difficulties in deducing information from quantum spectra which are interfered by numerous factors. Hence, the accuracy of the method depends steadily on the precision of reference approaches [10][29].

Ordinarily, UV-VIS-NIRS spectroscopic measurement is implemented in four standard modes: transmission, reflection, transflection (combine transmission and reflection), and interaction (similar to transflection but for solid samples) [29]. Measurement mode is selected depending on the type of sample and the location of installment.

3.2 Elastic light scattering

Elastic light scattering has been scientifically applied for characterising sizes and shapes of small particles for a long time [30][31]. It is also capable of distinguishing bacteria with different sizes and shapes in liquid suspensions [32][33]. Then, it can contribute to the turbidimetry and improve the result of bacterial estimation using a method called "scattering pattern" introduced by Shimizu *et al.* in 1978 [34]. The scattering pattern of bacteria is frequently clarified and modelled using Mie scattering [33][34]. In combination with Rayleigh-Gans-Debye approximation, it can determine the sizes and shapes distribution of bacteria [33].

Light scattering is the physical phenomenon happening when the trajectory of the light propagating through a medium behaves anomalously in multiple paths instead of a straight one. Due to that diffusion effect, light scattering attenuates the beam of light source traversing through the medium [30]. Scattering is typically categorised into three types. Firstly, Rayleigh scattering is applied when the size of the particle is smaller than the wavelength. Secondly, Mie scattering is utilised when the particle is larger than the wavelength. Finally, non-selective scattering is used when the particle is significantly larger than the wavelength.

In microorganism quantification, the scattering depends on the size and shape of microbial cells. The scattering problem is usually approximated with respect to a size parameter ν and a relative refractive index n of the transmitting medium [35]. The size parameter is defined as $2\pi r_p/\lambda$ where r_p is the dimension of the object and λ is the wavelength of the incident light. The refractive index of small soft particles similar to bacteria usually satisfies $\|n-1\| \ll 1$ [35]. Then, the scattering can be approximated with the aforementioned Rayleigh-Gans method. It is, however, limited to a small difference of the index of refraction between the particle and the surrounding environment.

3.3 Bacteria quantification

A large and growing amount of literature has investigated the application of spectroscopy on estimating the concentration of biological species. Typically, a particular wavelength is selected in most spectroscopic methods to measure the optical density (OD) of the molecules [25]. OD usually refers to the normalised value of the absorbance of the particles at a specific wavelength. That particular wavelength depends on the type and characteristics of the species. It is chosen at the absorption peaks, i.e. the strongest absorbed wavelengths, if they exist since they guarantee the robustness and sensitivity of the method [25]. The light attenuation is contributed by all molecules in a complex medium similar to the bacterial suspension which includes not only microbes but also biochemical substances.

In this work, UV-VIS spectroscopy was employed according to the standard and other literature. Multiple wavelength measurement allows a possibility to determine a suitable wavelength for bacterial characterisation with minor calibration. However, this spectroscopy is disadvantageous since it is incapable of classifying live and dead cells. Dead cells and cell debris also scatter light [25].

In this research, owing to high accuracy, NIRS was utilised in analysing solid and liquid samples with high signal-to-noise ratio (SNR) [10]. NIRS records the response of chemical bonds such as C–H, O–H, and N–H [10] which are present in the macromolecules of biological species (bacteria in this study) such as polysaccharides, lipids, nucleic acids, amino acids and proteins [36]. NIRS was applied to estimate the number of live cells in the suspension media (culturing media including bacteria).

The number of live cells, i.e. viability, is featured by total viable count (TVC) in biology. It is frequently measured with the plate count method through serial dilution. This approach was applied as the reference method in this project since it is one of the most acceptable methods in microbiology [36].

4 Chemometrics for spectroscopic data analysis and modelling

Data achieved from spectroscopic methods contains important, yet implicit information. Chemometrics is utilised in order to explicitly extract it. This chapter introduces the mathematical algorithms used for estimating and modelling bacteria in this project. The chapter contains five sections. The first section depicts the basic theory of chemometrics as a data analysis method for modelling the spectroscopic results. The second section describes some preprocessing methods for treating the spectral data in order to enhance significant information. The third and fourth sections present the principal component analysis (PCA) and the partial least square regression (PLSR) to construct the prediction model. The last section demonstrates a possible approach for optimising the prediction model.

4.1 Chemometrics

Chemometrics is defined, by Massart *et al.*, as "the chemical discipline that uses mathematical and statistical methods, (1) to design or select optimal measurement procedures and experiments, and (2) to provide maximum chemical information by analyzing chemical data" [37]. Chemometrics includes qualification and quantification to extract information from materials for developing classification and prediction models [10].

Chemometrics utilises mathematical and statistical methods of multivariate data analysis to extract the relation between samples such as food samples, drug samples, mineral, patients, or spectra, and measurement results such as nutritious elements, element constituents, concentration, absorbance, transmittance, reflectance, pH, or spectral peaks. Those methods can be categorised into five groups [38]:

- Multivariate regression methods for calibration model such as multiple linear regressions (MLR), multivariate calibrations (partial least squares regression – PLSR), and artificial neural network (ANN).
- Multivariate decomposition methods for reducing dimensions of data into correlated variables or components (latent variables) such as Principal component analysis (PCA) and Independent component analysis (ICA).
- Hierarchical cluster analysis to arrange components using their characteristics.
- Pattern recognition methods to classify data into confident regions such as Soft independent modeling of class analogy (SIMCA) and Linear discriminate analysis (LDA).

- Chemometric resolution methods to provide models illustrating the contribution of each component such as Generalized rank annihilation method (GRAM), Alternating least square (ALS), and Orthogonal projection analysis (OPA).

Chemometrics is applied in spectral analysis to elucidate the relations between multivariate spectral features, (e.g. absorption values at different wavelengths of samples), and the properties of the analytes, (e.g. the concentrations or physical attributes) [10][39]. Those features and properties form the independent and dependent variables, respectively. For the computational purpose, they are normally arranged into two matrices:

- Data matrix $\mathbf{X}$ is collected from samples of which each row $\mathbf{x}^{(i)}$ is an observation. Each column of $\mathbf{X}$ represents a feature and is typically known as an independent variable [10].
- Matrix $\mathbf{Y}$ holds the properties which are the dependent variables for the mathematical model. They are generally associated with an analytical reference method for the known samples and will be used to predict the properties of the unknown samples [10].

Chemometrics develops a model mathematically identifying a relationship between those types of variables as

$$\mathbf{Y} = f(\mathbf{X}) \tag{5}$$

where f is a relating function [10][40].

This project aims to construct a model identifying the relationship between bacteria viability (dependent variables) with respect to absorbance spectra (independent variables). Then, the work applies two chemometric methods including Principal component analysis (PCA) and Partial Least Square (PLS). PCA plays the role of decomposing the data matrix and reducing its dimensionality. After that, decomposed data was passed to Partial least squares (PLS) for modelling the relationship between spectra ($\mathbf{X}$) and bacterial viability ($\mathbf{Y}$). The chemometric tools are depicted in the following sections with respect to NIR spectral qualitative and quantitative analysis.

4.2 Measurement and data preprocessing

In chemometric process for spectroscopy, each observation is a spectrum; thus, a data matrix $\mathbf{X}$ is called a spectral data matrix. Each row $\mathbf{x}^{(i)} = (\mathrm{x}_j^{(i)}, \mathrm{x}_2^{(i)}, \ldots, \mathrm{x}_p^{(i)})$ consists of measured data at p different variables equivalent to p wavelengths. Each element $\mathrm{x}_j^{(i)}$ contains a measured value such as intensity, absorbance, or transmittance at a wavelength [nm], or a wavenumber [cm^{-1}]. N spectra are collected into row vectors $\mathbf{x}^{(i)}$ (i = 1, 2, ..., N) of the data matrix $\mathbf{X} \in \mathbb{R}^{N \times p}$ as below:

$$\mathbf{X} = \begin{bmatrix} \mathrm{x}_1^{(1)} & \mathrm{x}_2^{(1)} & \mathrm{x}_3^{(1)} & \dots & \mathrm{x}_p^{(1)} \\ \mathrm{x}_1^{(2)} & \mathrm{x}_2^{(2)} & \mathrm{x}_3^{(2)} & \dots & \mathrm{x}_p^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{x}_1^{(N)} & \mathrm{x}_2^{(N)} & \mathrm{x}_3^{(N)} & \dots & \mathrm{x}_p^{(N)} \end{bmatrix}$$

In chemometrics, data preprocessing can partly eliminate undesired variants from collected data and enhance the signal-to-noise ratio (SNR) for the efficacy of data processing and quality analysis. Those variants result typically from differences in size, morphology (e.g. shape and roughness of sample surface), or geometry (e.g. the position of the optical equipment with respect to objects), or density variations (e.g. the inhomogeneous distribution of particles). A vast of pretreatment methods are commonly used such as baseline correction, mean-centering, de-trending, smoothing, normalisation, orthogonal signal correction, standard normal variate transformation, multiplication scatter correction and Savitsky-Golay derivation[10].

This project utilised the mean-centering and the smoothing filter. In particular, the values of each measurement are tailored with mean-centering through two steps:

1. calculate the mean value for each feature $\eta_j = (1/N)\sum_{i=1}^{N} \mathrm{x}_j^{(i)}$ for j from 1 to p
2. subtract mean value from each corresponding value $\mathrm{x}_j^{(i)} = \mathrm{x}_j^{(i)} - \mu_j$ ($i = 1, 2, \dots, N$; $j = 1, 2, \dots, p$)

Smoothing filter is another approach applied in this project. It eliminates minor fluctuation which contributes uselessly to the data set. Especially, it smooths the noisy spectra obtained from chemical analysers. Basically, the filter replaces each value f_j corresponding to a point x_j in a series of points with a value obtained as the average of 2n + 1 points, where n is the number of left and right neighbours of x_j [41]. The new value g_j is computed as

$$g_j = \frac{\sum_{k=j-n}^{k=j+n} x_k}{2n+1}. \tag{6}$$

4.3 Principal component analysis

Principal component analysis (PCA) is commonly utilised in finding patterns in high-dimensional data [42]. From the mathematical point of view, it is a simple orthogonal transformation method for compressing the data dimensionality. Hence, it is extensively applied in multivariate data analysis and other fields such as face recognition and image compression [42]. The original data set is ordinarily formed from a substantial number of correlated variables. After transforming, it can be represented with a significantly smaller set of uncorrelated variables (latent variables) named principal components (PCs), which still maintain the most important information [10][43].

PCA deduces a new set of projective directions along which the data varies the most. They are computed so that the first PC describes the most variance of the data [39]. The second PC spans in the direction that contains maximum variance subject to being uncorrelated with the first one. Simply, it means that the second component is orthogonal to the first component [40]. A simple illustration of PCA in two-dimensional space can be seen in Figure 2.

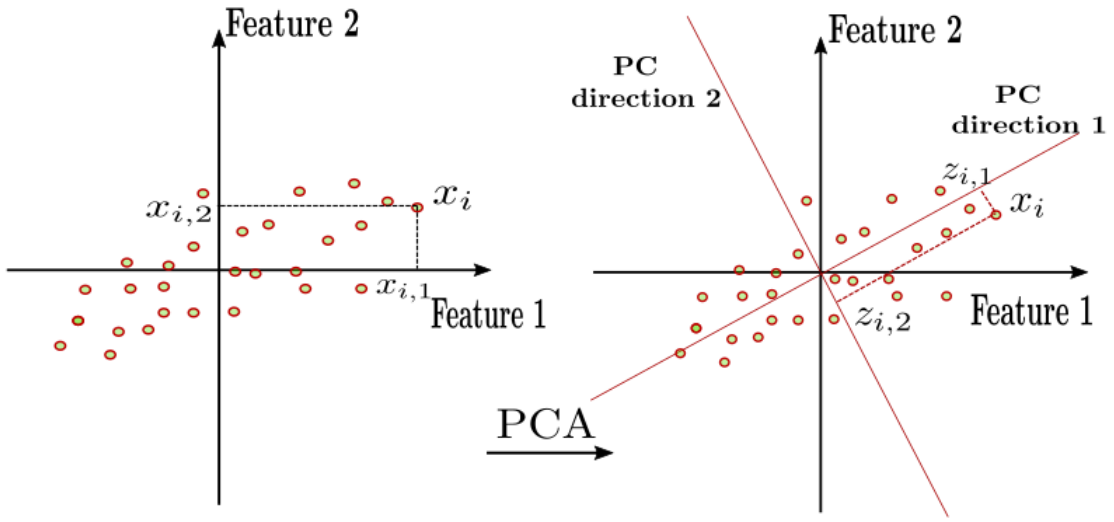


Figure 2: Example of two first principal components in 2-D space.

Let $\mathbf{P} \in \mathbb{R}^{p \times m}$ be the projection matrix of which columns are the directional vectors of PCs. Those vector are orthogonal to each other. The data matrix $\mathbf{X}$ can be projected onto those PCs and presented as

$$\mathbf{T} = \mathbf{XP} \tag{7}$$

where $\mathbf{T} \in \mathbb{R}^{N \times m}$ is the new data matrix. At most, p components can be achieved when all the features from the original data set are linearly independent. However, due to covariant manner of natural data sets, the number of PCs is typically much less than the number of features, i.e. $m \ll p$ [43]. In summary, after computing with PCA, two matrices can be obtained:

- Matrix $\mathbf{P}$ in equation (7) includes the orthogonal vectors of PCs. Each of them has an equal dimension to the original number of variables. This matrix is also known as the matrix of *Loadings*. It demonstrate both the variation of the data and the level of importance of each original variable to the data set [39].
- Matrix $\mathbf{T}$ in equation (7) is the new data set and known as the matrix of *Scores*. Each column of T shows the variation of each sample with respect to each PC [39]. Plotting Scores can support interpreting data in the PCA, for instance, identifying outliers [10].

The illustration of Loadings and Scores in spectral analysis can be found later in Figure 10 and Figure 11 in Subsection 5.2.2.

Implementation

PCs can be simply determined using (1) eigenvalues and eigenvectors of covariance matrix [44], (2) singular value decomposition (SVD) from the data matrix equivalently [44], or (3) nonlinear iterative partial least squares (NIPALS) method [40] which can advance further for computing PLS. Those methods can be explained as follows:

- ***Eigenvalues and eigenvectors of covariance matrix***
  Firstly, data set needs to be projected onto uncorrelated directions [44]. Relationship between pairs of measurements in the data set is presented with covariance matrix and is computed with the below equation [44]:

$$\mathbf{S_X} = \frac{1}{N-1}\mathbf{X'X} \tag{8}$$

  Completely uncorrelated data yields zero-covariance; whereas highly correlated data results in high covariance meaning that they record the same dynamics information of the system, which means redundancy [44]. The redundancy should be mininised by choosing new PCs [44].

  Firstly, we are interested in determine all PCs. To this end, let $\mathbf{P_t} \in \mathbb{R}^{p\times p}$ denote the matrix containing all PCs. Then $\mathbf{P_t}$ is the transformation that transforming $\mathbf{X}$ to a new data matrix $\mathbf{H} = \mathbf{XP_t} \in \mathbb{R}^{N\times p}$ [44]. The covariance matrix of $\mathbf{H}$ can be derived:

$$\begin{aligned}\mathbf{S_H} = \frac{1}{N-1}\mathbf{H'H} &= \frac{1}{N-1}(\mathbf{XP_t})'(\mathbf{XP_t}) \\ &= \frac{1}{N-1}\mathbf{P_t'X'XP_t} = \mathbf{P_t'S_XP_t}.\end{aligned} \tag{9}$$

  Then, this method seeks an orthogonal matrix $\mathbf{P_t}$ to diagonalise the covariance $\mathbf{S_H}$. When choosing column vectors $\mathbf{p}_i$ of $\mathbf{P_t}$ as eigenvectors of $\mathbf{S_X}$, $\mathbf{P_t}$ is orthonormal and $\mathbf{S_H}$ is diagonalised as proved by Shlens (2003) [44]:

$$\mathbf{S_H} = \mathbf{P_t'S_XP_t} = \mathbf{P_t'}(\mathbf{P_t\Gamma P_t'})\mathbf{P_t} = \mathbf{\Gamma} \tag{10}$$

  where $\mathbf{\Gamma}$ is a diagonal matrix of which diagonal elements correspond to the eigenvalues of $\mathbf{X'X}$. Therein, the diagonal elements of $\mathbf{S_H}$ show the variance of $\mathbf{X}$ along $\mathbf{p}_h$.

  Secondly, PCs should be chosen so that they can represent the data with a minimised amount of noise. Normally, PCs with larger associated variances depict compelling dynamics of the system whereas the others with lower variances represent noise [44]. The first component direction corresponds to the largest eigenvalue, σ_1^2 of $\mathbf{X'X}$. The second component direction corresponds to the second largest eigenvalue, σ_2^2. Similar result holds for other components [44][42]. Then, the PCs associated with higher elements of $\mathbf{S_H}$ are selected as the new PCs forming matrix $\mathbf{P}$.

Therefore, this method is performed through four basic steps [42]:

1. Mean-center data to get matrix $\mathbf{X}$
2. Calculate the covariance matrix $\mathbf{S_X}$ in equation (8).
3. Calculate the eigenvectors and eigenvalues of the covariance matrix and arrange them in descending order to achieve matrix $\mathbf{P}$.
4. Choose new components and form featuring vectors to form matrix $\mathbf{P}$. The number of new components is usually chosen arbitrarily but they should normally cover 95 – 99 % of the original data set. Choosing too many of them can lead to over-fitting in the later model, i.e. the model can include noise and other interferences. Too few chosen components can result in underfitting or loosing important information [42].

- ***Singular Value Decomposition***

  SVD is an algorithm to find the solution for PCA. It is performed through basic steps like the covariance method but replacing step 2 and 3 with SVD. A matrix $\mathbf{X}$ can be decomposed into $\mathbf{X} = \mathbf{D\Sigma V'}$.
  - $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N)$ is a $N \times N$ orthogonal matrix.
  - $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$is a $p \times p$ orthogonal matrix.
  - $\mathbf{\Sigma} \in \mathbb{R}^{N \times p}$ is diagonal matrix including singular values in descending order.

  Then, we have

$$\begin{aligned} \mathbf{X'X} = (\mathbf{D\Sigma V'})'(\mathbf{D\Sigma V'}) &= \mathbf{V\Sigma' D' D\Sigma V'} \\ &= \mathbf{V\Sigma' \Sigma V'} = \mathbf{V\Sigma^2 V'} \end{aligned} \tag{11}$$

  Therefore, the columns of $\mathbf{V}$ are also the eigenvectors of the covariance matrix $\mathbf{X'X}$. They are analogous to the PCs of $\mathbf{X}$ found in the covariance method. Hence, the columns of $\mathbf{P}$ can be determined by the m eigenvectors corresponding to the m largest eigenvalues of $\mathbf{X}$. The matrix $\mathbf{\Sigma}^2$ is a diagonal matrix including the eigenvalues of $\mathbf{X}$.

- ***Nonlinear iterative partial least squares (NIPALS)***

  Nonlinear iterative partial least square(NIPALS) is another method used for calculating PCs. Matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1' + \mathbf{t}_2\mathbf{p}_2' + ... + \mathbf{t}_m\mathbf{p}_m' = \mathbf{TP'} \tag{12}$$

  where $\mathbf{t}_h \in \mathbb{R}^{N \times 1}$ is a score vector and $\mathbf{p}_h \in \mathbb{R}^{p \times 1}$ is a loading vector. NIPALS starts with calculating $\mathbf{t}_1$ and $\mathbf{p}_1'$ from the $\mathbf{X}$ matrix. Then the outer product, $\mathbf{t}_1\mathbf{p}_1' \in \mathbb{R}^{N \times p}$, is subtracted from $\mathbf{X}$ to yield the residual $\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1'$. This residual is used for calculating $\mathbf{t}_2$ and $\mathbf{p}_2'$. The process is repeated until the last PC is found. The algorithm is implemented as follows [40]:

Algorithm 1: NIPALS algorithm

Input: $\mathbf{X} \in \mathbb{R}^{N \times p}$ data matrix of samples
Output: Principal components, matrix of loadings $\mathbf{T} \in \mathbb{R}^{N \times m}$ and matrix of scores $\mathbf{P} \in \mathbb{R}^{p \times m}$

`/* Take a vector from X */`
$\mathbf{t}_h = \mathbf{x}_j$;
while $\mathbf{t}_{h\ new} \neq \mathbf{t}_h$ **do**
  $\mathbf{p}'_h = \mathbf{t}'_h\mathbf{X}/\mathbf{t}'_h\mathbf{t}_h$;
  `/* Normalise */`
  $\mathbf{p}'_{h\ new} = \mathbf{p}'_{h\ old}/||\mathbf{p}'_{h\ old}||$;
  $\mathbf{t}_{h\ new} = \mathbf{X}\mathbf{p}_h/\mathbf{p}'_h\mathbf{p}_h$

When NIPALS converges, the results have been proved to be the same with ones achieved by using eigenvectors [40].

Spectroscopic data are obtained from up to hundreds of samples – each sample is relevant to a spectrum – and each spectrum consists of hundreds to thousands of wavelengths. A large number of them are correlated, meaning that the measured absorbances (or other attributes) at different wavelengths are dependent on others. This dependence leads to weak prediction performance in the calibration model to be developed. Hence, PCA is used for compressing the data set and getting a better comprehension through uncorrelated data.

After processing with PCA, the Scores and the Loadings of spectra can be obtained and plotted. Each sample will have a specific score over each PC. A scatter plot of the score on one PC versus score on another PC of each sample can cluster the samples into distinguishable groups. Plotting of loadings shows the weight of each wavelength on the four first PCs. PCA helps estimate how much each of the original peaks contributes to each new PCs.

In spectral analysis, principal components usually indicate the structure of particles contained in the samples which influence the variation of the spectra [43]. Hence, PCs extracted from PCA are typically used as input for a classifier to discriminate the content of each sample. The similar principle is applied in biology. Therein, PCA is regularly applied as pre-processing step for classifying the growth phases of bacteria [45][46] or segregating the types of bacteria in the samples [47][48].

4.4 Partial least squares regression

Partial least squares (PLS) is a multivariate data analysis method introduced in the late sixties by the Swedish mathematician Herman Wold for serving social science. Then the method was improved in cooperation with his son in the late seventies for chemical applications [40]. Nowadays, PLS is one of the most popular linear regression methods for analysing spectroscopic data in chemical engineering, clinical chemistry, industrial process control, bioinformatics, neuroscience, and other fields [10]. In chemistry, PLS is normally utilised for predicting the concentrations of constituents contributing to the ingredients of compounds. In microbiology, it is a frequent approach to estimate the concentration of bacteria in liquids (e.g. milk and juice) or the density of bacteria on solid food (e.g. pork, lamb, and fish) [11][46][49][50].

PLS theoretically consists of PCA and is considered as the next step to extract the model from chemometric process. PCA extracts PCs as the sources of variation from data and captures only the characteristics of independent variables in $\mathbf{X}$. On the other hand, PLS correlates them with dependent variables in $\mathbf{Y}$ (e.g. concentration data) of known samples to infer Latent Variables (LVs)[39]. In other words, PCA produces a weight matrix demonstrating the covariance between independent variables while PLS computes weights reflecting the covariance between those independent variables and the dependent variables. PLS can be used for modelling as in equation (5). The model, then, is applied for quantitatively predicting $\mathbf{Y}$ values for unknown samples. The general procedure of PLS data analysis is described in Figure 3 [40].

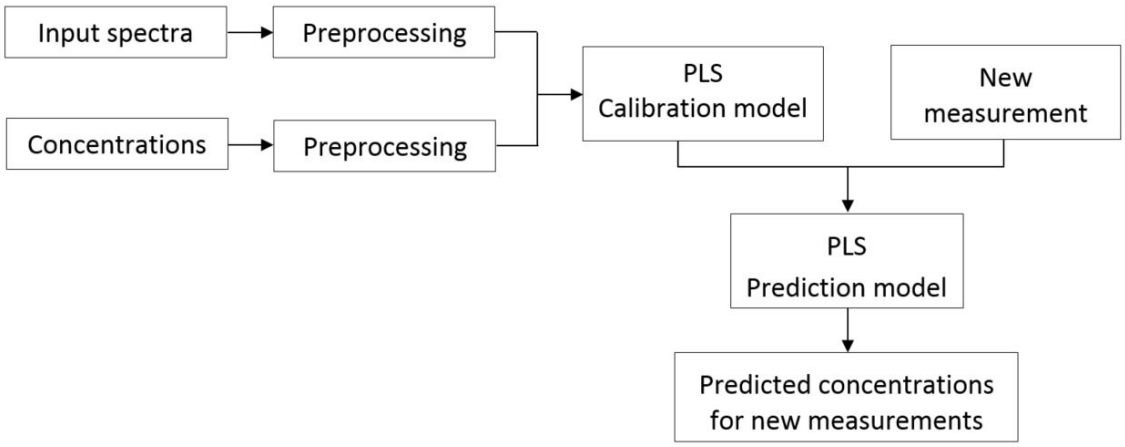


Figure 3: Constructing calibration model process [39].

Calibration model

Partial least square (PLS) is an extension to multiple linear regression (MLR). MLR model simply illustrates the linear relationship between a set of dependent variables in $\mathbf{Y}$ and independent variables in $\mathbf{X}$. In spectroscopic analysis, PLSR computes a

calibration model in the same sense with the form:

$$\mathbf{Y} = \mathbf{XB} \tag{13}$$

where $\mathbf{X} \in \mathbb{R}^{N \times p}$ is spectra matrix, $\mathbf{Y} \in \mathbb{R}^{N \times q}$ is concentration matrix, $\mathbf{B} \in \mathbb{R}^{p \times q}$ is the regression coefficient matrix or produced model, and q is the number of dependent variables [39].

In order to build a calibration model, relevant variation should be retained and irrelevant one should be eliminated as much as possible. The obtained model maximises the covariance between $\mathbf{X}$ and $\mathbf{Y}$. The data used for constructing the model is called calibration or training set [40]. The quality of the calibration model is determined with Root Mean Square Error in Calibration (RMSEC). RMSEC quantifies the differences between the predicted values from the model and the realistically true values of the training set (i.e training error). Hence, the smaller RMSEC is, the better the model fits the data. Typically, RMSEC is chosen within a suitable range depending on the scope and requirements of each project.

Several methods can be applied to select appropriate calibration data since it plays a decisive role in the final prediction model. Firstly, Design of Experiments (DoE) is utilised to effectively map a data space. It uses optimal methods such as D-optimal, E-optimal, and Kennard-Stone methods. Secondly, sample sets are selected randomly, analysis is performed, and the errors are calculated. This whole process is repeated several times before the average errors are determined at the end.

Implementation

In a distinguishable manner compared to PCA, PLS computes individually outer relations for each block $\mathbf{X}$ and $\mathbf{Y}$ and inner relation linking between the two blocks. The outer relations are presented as:

$$\begin{aligned} \mathbf{X} &= \mathbf{TP}' + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ}' + \mathbf{F} \end{aligned} \tag{14}$$

where $\mathbf{T} \in \mathbb{R}^{N \times m}$ and $\mathbf{U} \in \mathbb{R}^{N \times r}$ are score matrices, $\mathbf{P} \in \mathbb{R}^{p \times m}$ and $\mathbf{Q} \in \mathbb{R}^{q \times r}$ are loading matrices, $\mathbf{E} \in \mathbb{R}^{N \times p}$ and $\mathbf{F} \in \mathbb{R}^{N \times q}$ are noise terms for $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then, the inner relation is presented as:

$$\hat{\mathbf{u}}_h = b_h \mathbf{t}_h \tag{15}$$

where $\hat{\mathbf{u}}_h = \mathbf{u}_h' \mathbf{t}_h / \mathbf{t}_h' \mathbf{t}_h$, $\mathbf{u}_h$ is a column vector of the h-th component of $\mathbf{U}$, b_h is the regression coefficient for the h-th component, $\mathbf{t}_h \in \mathbb{R}^{N \times 1}$ is a column vector of of the h-th component of $\mathbf{T}$. Several algorithms can be utilised to compute PLSR model. One of them is NIPALS which has been proved to be more robust and accurate than others. It can be implemented as follows [40]:

Algorithm 2: Partial least squares using NIPALS algorithm

Input: $\mathbf{X} \in \mathbb{R}^{N \times p}$ data matrix of samples,
$\mathbf{Y} \in \mathbb{R}^{N \times q}$ concentration of known samples,
$\mathbf{X}_u \in \mathbb{R}^{r \times p}$ data matrix of unknown samples,
ncomp as the number of components
Output: $\mathbf{Y}_u \in \mathbb{R}^{r \times q}$ concentration of unknown samples

$\mathbf{E} = \mathbf{X}$; $\mathbf{F} = \mathbf{Y}$;
$\mathbf{t}_{old} = 100000$; /* An arbitrary large starting value */
$\mathbf{t} = 90000$; /* An arbitrary starting value */
$\epsilon = 1e-6$; /* A threshold to stop the algorithm */
for h = 1 : ncomp **do**
  $\mathbf{u}_{start} =$ some $\mathbf{y}_j$; /* take a vector from $\mathbf{Y}$ */
  while $\|\mathbf{t} - \mathbf{t}_{old}\| > \epsilon$ **do**
    $\mathbf{t}_{old} = \mathbf{t}$;
    /* For X */
    $\mathbf{w}' = \mathbf{u}'\mathbf{E}/\mathbf{u}'\mathbf{u}$;
    $\mathbf{w}' = \mathbf{w}'/||\mathbf{w}'||$; /* Normalisation */
    $\mathbf{t} = \mathbf{E}\mathbf{w}/\mathbf{w}'\mathbf{w}$;
    /* For Y */
    $\mathbf{q}' = \mathbf{t}'\mathbf{F}/\mathbf{t}'\mathbf{t}$;
    $\mathbf{q}' = \mathbf{q}'/||\mathbf{q}'||$; /* Normalisation */
    $\mathbf{u} = \mathbf{F}\mathbf{q}/\mathbf{q}'\mathbf{q}$;
  $\mathbf{p}' = \mathbf{t}'\mathbf{E}/\mathbf{t}'\mathbf{t}$;
  /* Collect vectors for the h-component */
  $\mathbf{q}_h = \mathbf{q}$; /* $\mathbf{q}_h$ is a column vector of $\mathbf{Q}$ */
  $\mathbf{u}_h = \mathbf{u}$; /* $\mathbf{u}_h$ is a column vector of $\mathbf{U}$ */
  $\mathbf{t}_h = \mathbf{t}||\mathbf{p}'||$; /* $\mathbf{t}_h$ is column vector of $\mathbf{T}$ */
  $\mathbf{w}_h = \mathbf{w}||\mathbf{p}'||$;
  $\mathbf{p}_h = \mathbf{p}/||\mathbf{p}||$; /* $\mathbf{p}_h$ is a column vector of $\mathbf{P}$ */
  /* Find regression coefficients */
  $b_h = \mathbf{u}'_h\mathbf{t}_h/\mathbf{t}'_h\mathbf{t}_h$; /* b_h is a regression coefficient */
  /* Residuals $\mathbf{E}$ and $\mathbf{F}$ are calculated as: */
  $\mathbf{E} = \mathbf{E} - \mathbf{t}_h\mathbf{p}'_h$;
  $\mathbf{F} = \mathbf{F} - b_h\mathbf{t}_h\mathbf{q}'_h$;
/* Compute the unknown concentrations */
$\mathbf{E} = \mathbf{X}_u$; $\mathbf{Y}_u = 0$;
for h = 1 : ncomp **do**
  $\mathbf{t} = \mathbf{E}\mathbf{w}_h$;
  $\mathbf{E} = \mathbf{E} - \mathbf{t}\mathbf{p}'_h$;
  $\mathbf{Y}_u = \mathbf{Y}_u + \mathbf{b}_h\mathbf{t}\mathbf{q}'_h$

Testing and evaluating

Validation data is applied for checking the predictive performance of the model. Validation can be performed using the subsets of the calibration data as Cross Validation or newly collected data in True Validation [39].

One method of validation is cross validation. It can assess the predictive ability of a potential model and is usually used when the number of samples is small. After that, a number of reasonable components can be chosen to retain the model. In order to implement cross validation, the data set is divided into several partitions. One or some of them are utilised to build a model. They are known as training sets. Then one partition is used for validating the model. It is known as testing set. Typically, there are three methods can be used for cross validation [51]:

- ***Holdout*** method is performed by separating the data set into two smaller sets called training set (regularly it consists of 70–75 % of data set) and testing set (the other 25–30 % of data set). From there, the training set is utilised to fit a model, and then the other set is used for validating the model. Errors are accumulated after calculation to evaluate the model. Despite the advantage of computing time, the results are not well explained because the division of the data set can result in the variation of the model.
- ***K-fold*** method is implemented by splitting the original data set into k subsets. The algorithm runs iteratively k times. Each time one subset is used as the testing set and the other k – 1 ones are used as training sets. The average errors are computed after running k times. Each subset becomes training set k – 1 times and testing set 1 time. It overcomes the problem of Holdout method, especially when k is large. However, a drawback of this method is laborious computation since it repeats k times. A high value of k requires extensive computational resources.
- ***Leave-one-out*** method is carried out similar to K-fold method but only one of the data point is left out for evaluation each time. It means that if the data set includes N observations, N – 1 observations are applied for modelling. One observation is used for evaluating and computing the errors. At a first glance, it seems to be expensive to compute by the K-fold method. However, computing the error for one observation is only as easy as computing the residual error. Hence, it costs much less than the K-fold method.

Quality measurement of validation is obtained with Root Mean Square Error of Cross Validation (RMSECV) or Root Mean Square Error of Prediction (RMSEP) or both. RMSECV is used when the number of observations is significantly small compared to the number of variables. RMSEP evaluates the differences between the predicted values from the model and the realistically observed values of the new measurement.

In general, the RMSE is calculated as [52]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2}{N}} \tag{16}$$

where $\hat{y}^{(i)}$ is the estimated value of the analysed feature, $y^{(i)}$ is the measured value of the analysed feature, and N is the number of samples in the particular data set. Statistically, the low RMSECV and RMSEP are preferred when choosing the model [49] [50].

In the statistical analysis, a single standard is frequently insufficient for evaluating a model because each of them retains some limitation [52]. As a result, several coefficients are frequently utilised together for evaluating the precision and prediction capability of the model. Another popular standard is the coefficient of determination R^2. The higher value of R^2 is normally interpreted as a better model [49] [52]. A model with $R^2 > 0.9$ is usually considered as a very good model. Then, the best model is selected for each attribute based on highest R^2 and lowest RMSEC and RMSECV.

Number of latent variables

The number of LVs for establishing the model affects the quality of calibration and validation significantly [39][40]. Too few LVs will generate ill prediction because of inadequate information. Too many LVs will lead to over-fitting model because the robustness is lost to a small amount of variation. Additionally, the measured data is never free of noise and some components carry only noise. Besides, some other components generate a problem of collinearity and should be excluded [40][49]. Consequently, some components are commonly left out for a high predictive power.

Several means can be adopted to find an adequate number of LVs. One possibility is choosing the number of LVs by either determining the minimum value of predicted residual error sum of squares (PRESS) in equation (17) or root mean squared error cross validation in equation (18) [40][53]:

$$PRESS = \sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2 \tag{17}$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2}{N}} \tag{18}$$

where $\hat{y}_i$ is the estimated value, y_i is the measured value and N is the number of samples in the particular data set. Another method is to compare RMSEC and RMSEP and to select the number of LVs at which they are closest to each other, as shown in Figure 4. Additionally, model needs to be maintained regularly by collecting new lab data over time and re-validating the model [39].

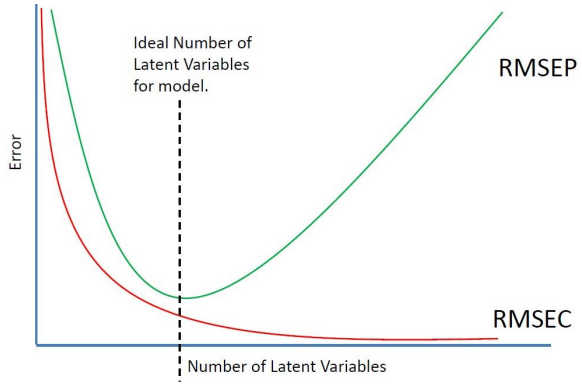


Figure 4: Choosing latent variables [39].

4.5 WRC-PLSR for optimal wavelengths selection

Using the full spectrum with hundreds of wavelengths enables a high possibility of capturing the wavelengths that contain feature information. However, using a full scan not only implies a possibility of an overfitting model but also requires a high computational capacity for data analysis, which should be prevented in industrial applications [50] [53]. Besides, optimal wavelengths may represent information equally or more efficiently than the full spectrum. In this case, proper selection can be conducted for the wavelengths emphasising most of the characteristic information of the spectra [52]. Therefore, a number of studies have been devoted to improve PLS model by choosing an appropriate number of wavelengths. Different methods have been proposed such as intermediate least square regression (ILS), interactive variable selection (IVS) for PLS, weighted regression coefficients (WRC) or genetic algorithm (GA) [54] [55]. In this project, PLS-WRC was utilised for selecting the optimal wavelengths.

WRC-PLSR was introduced by Frenich, *et al.* (1997) [53] and is constructed similarly to PLS model. The coefficients b from matrix $\mathbf{B}$ of PLS model cannot be applied directly to choose optimal wavelengths. It is because a large b can represent either a significant variable or a variable with a small absolute value but a large variance [49][53]. The problem can be prevented by standardising the input data. This is implemented by weighting the variables in $\mathbf{X}$ with the inversed standard deviation [53]:

$$s_j = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{x}_j^{(i)} - \eta_j)^2}{N-1}} \tag{19}$$

where s_j is the standard deviation of variable x_j, $\{\mathrm{x}_j^{(1)}, \mathrm{x}_j^{(2)}, \mathrm{x}_j^{(3)} ..., \mathrm{x}_j^{(N)}\}$ are the observed values of samples of the variable x_j, η_j is the mean value of those observations, and N is the number of samples. After the process, each variable x_j has the same variance. The problem (13) becomes:

$$\mathbf{Y} = \mathbf{X}_{standadised}\mathbf{B}_w + \mathbf{F} \tag{20}$$

where $\mathbf{B}_w \in \mathbb{R}^{p\times q}$ is the new regression coefficient matrix. Then, the coefficients β in $\mathbf{B}_w$ with large absolute values indicate informative wavelengths. In order to compute the new weighted matrix, PLS algorithm determines the orthogonal projection axes W or PLS-weights with the relation [53]:

$$\begin{aligned} \mathbf{T} &= \mathbf{XW} \\ \mathbf{B}_w &= \mathbf{W}(\mathbf{P}'\mathbf{W})^{-\mathbf{1}}\mathbf{Q}' \end{aligned} \tag{21}$$

where $\mathbf{W} \in \mathbb{R}^{N\times m}$ is the weight loadings of $\mathbf{X}$, $\mathbf{P}$ is the loadings matrix of $\mathbf{X}$ and $\mathbf{Q}$ is the loading matrix of $\mathbf{Y}$. The whole procedure undergoes the following steps [53]:

1. Achieving an optimum model and weighted-regression coefficients matrix $\mathbf{B}_w$ using standardised data.
2. Choosing LVs using predictive error RMSECV from leave-one-out cross validation as in equation (18).
3. Plotting the β coefficients of the optimum model and selecting the optimum wavelengths as the peaks higher than a threshold.
4. Developing a new model using chosen wavelengths.
5. Reevaluating the new model with leave-one-out cross validation.

This chapter has introduced the analytical tools used to extract the data and to construct an optimal model. The next chapter will introduce the procedure to prepare and implement experiments.

5 Phantom experiments

This chapter presents the necessary materials and methods implemented in this project as well as the two phantom experiments. The phantom experiments utilised LEDs and colour liquids as the objectives of measurements. LED lights and colour liquids maintained their conditions over time. Hence, they provided controlled conditions to verify the utility of the methods. This chapter is divided into three sections. The first section provides basic information about optical equipment used in the project. The next two sections are presented in accordance with two phantoms executed throughout the project. The first phantom aimed to measure the radiance of a group of LEDs and to analyse the information encompassing in their spectra. The second phantom was arranged for constructing a model which can predict the concentrations of different food colour liquid. The second and third sections provide information about setting up the system, calibrating the spectrometer, collecting the spectra and examining the result.

5.1 Optical equipment

The following optical apparatuses were utilised during the whole research:

- A spectrophotometer (Ocean Optics HR4000) of which spectrum spans from 195 to 1118 nm [56]. This range encloses the spectra of LEDs emitting visible light. It was employed to acquire spectral data.
- An integrating sphere (Ocean Optics ISP-REF). It collects the power of incident radiants from all directions.
- SpectraSuite software (Ocean Optics 2007) for retrieving spectral data.
- A halogen light source providing wavelength range from 360–1700 nm (Ocean Optics Mikropack HL-2000-FHSA). This is a blackbody with the bulk temperature of 2500 K [57]. It played the role as a calibration source for relative irradiance as well as a radiant source for absorbance measurement in a later stage.
- Two 600-µm optical fibres (Thorlabs, USA), one for transmitting light from the Halogen light source to the system for calibration and the other for transmitting signal from the integrating sphere to the spectrometer for measurement.

Some other devices were also used depending on the objectives of experiments.

5.2 Phantom with LEDs

In the first phantom, the radiance spectra of four LEDs were acquired using the optical apparatuses mentioned in Section 5.1. The LEDs were used including blue, yellow, red and white os which typical spectra are shown in Figure 5. Additionally, an Arduino micro-controller was used to provide the signal to the system, four potentiometers were utilised to control the power of the LEDs. The system was constructed as demonstrated in Figure 6. A fibre was connected between the spectrophotometer and the integrating sphere through SMA (SubMiniature version A) connectors. All movement was prevented by using tapes to maintain the cables on the table during calibrating and measuring. Therein, the disturbances and errors could be minimised in the results.

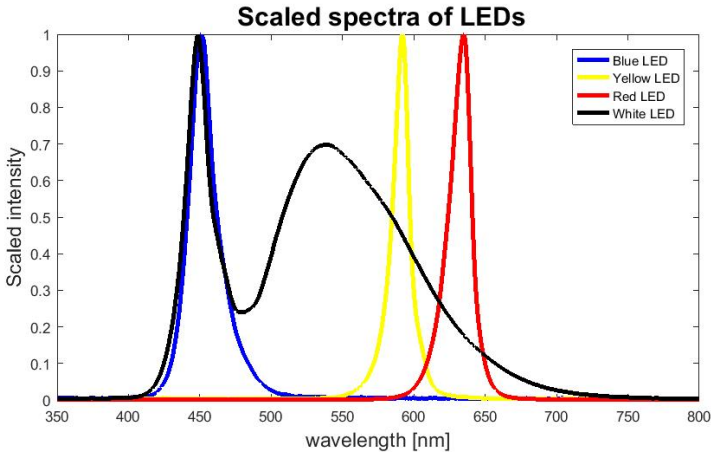


Figure 5: Normalised spectra of the LEDs used in the first phantom.

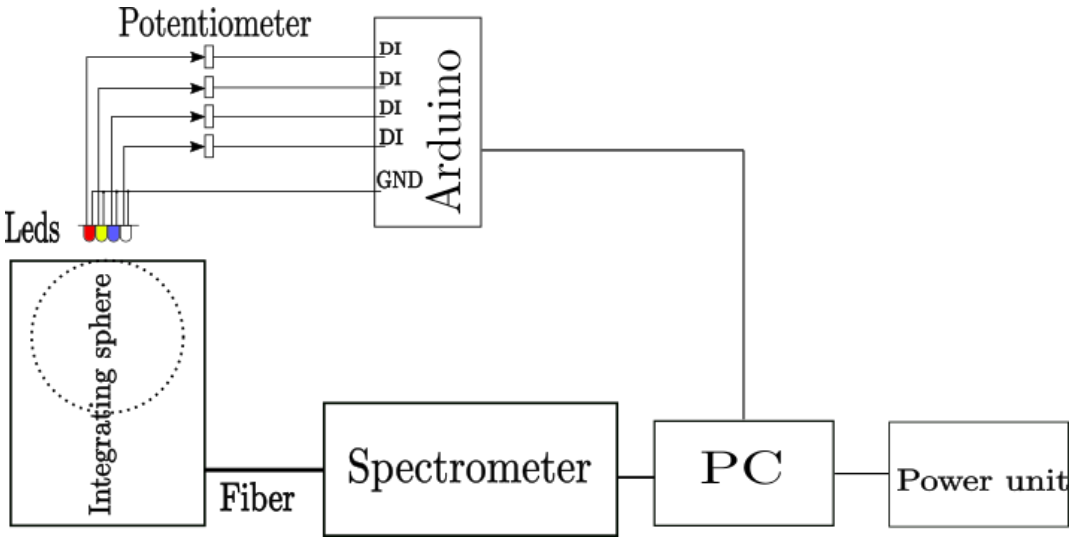
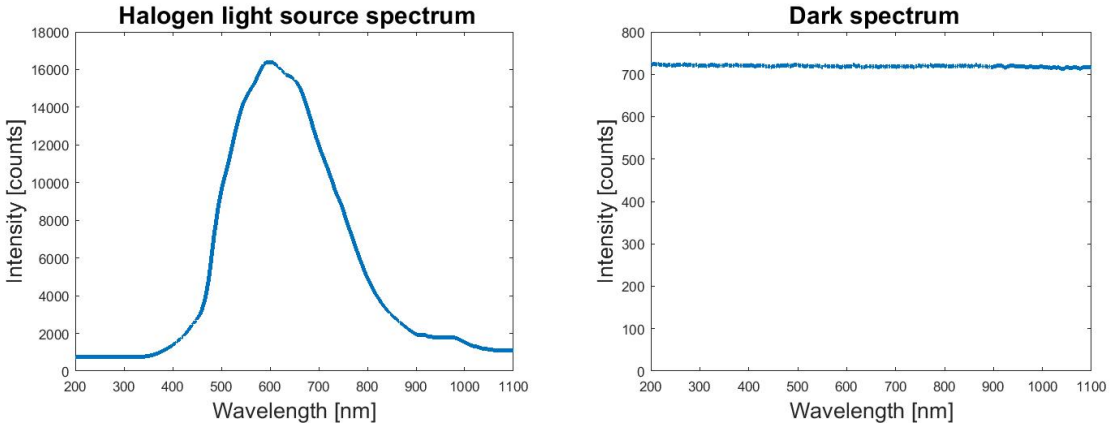


Figure 6: Phantom system set up.

5.2.1 Spectral acquisition for LEDs

Prior to every measurement, the optical system required warm-up and calibration. The halogen light source needed warming up for at least 20–30 minutes in order to achieve its thermal stability [57]. The same amount of time was also dedicated to warm up the spectrophotometer [56]. Before implementing measurement, a calibration process was performed with the light source and the integrating sphere. The shutter of the light source was adjusted in order to provide enough light to the spectrophotometer and to allow achievable integration time (i.e. the time interval that the detector collects photons before passing the measured values to A/D converter). The integration time was acquired automatically using the Relative irradiance mode from SpectraSuite software. It was set to approximately 19 ms as the intensity of the peak at 600 nm of the halogen light source reached the highest possible value, approximately 15000 counts, but was not saturated. A reference spectrum from the halogen source was captured and stored. Then, a dark spectrum (thermal noise) was obtained by closing the shutter so that no light transversed into the spectrometer. The obtained reference and dark spectra are illustrated in Figure 7.



(a) (b)

Figure 7: (a) The reference spectrum of the halogen light source and (b) the dark spectrum obtained in phantom experiment.

In each measurement, different combinations of LEDs were arranged by increasing or decreasing the resistance values of the potentiometers. The highest resistance value corresponded to the dimmest level of the LEDs and the lowest resistance allowed the brightest level. Each combination was considered as a sample. The distinct intensity spectra were collected with High speed acquisition mode of SpectraSuite. Fifty spectra of each combination could be achieved over a short period of time due to low integration time as shown in Figure 8.

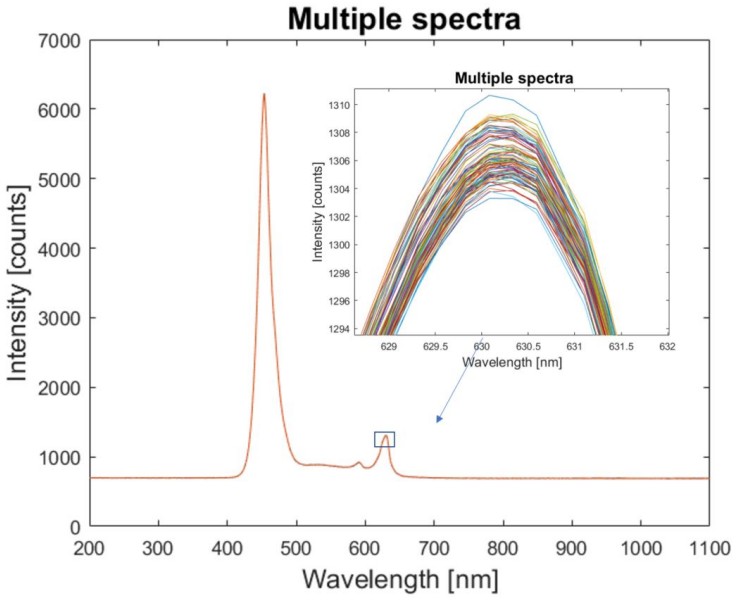
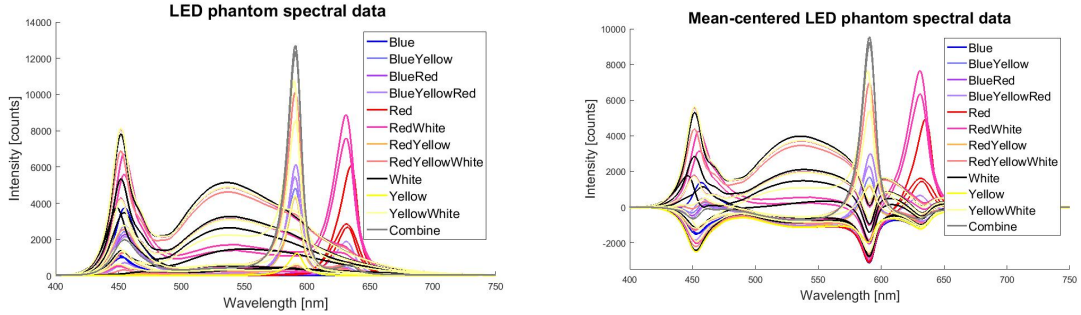


Figure 8: Multiple spectra of a group of blue and red LEDs.

5.2.2 PCA result for LEDs

A total of 40 spectra were measured in the project. Each measured spectrum included 3648 measurement points obtained in the range from 195–1118 nm with an interval of approximately 0.25 nm. This number of points was unnecessarily high to the analysis. Hence, the spectra were down-sampled to 912 points at 1 nm step without losing important features. Multiple spectra of each sample were averaged to achieve a single spectrum representing that specific sample. Then, the spectra were mean centered as described in Section 4.2.

All the statistical analyses in this project were performed in MATLAB 2016b (Mathworks). The input '.ProcSpec' files were imported to MATLAB workspace using function importoceanoptics [58]. The achieved spectra (Figure 9a) were averaged and mean centered as can be seen from Figure 9b. They were subjected to pca function of MATLAB for extracting principal components.



(a) (b)

Figure 9: LED phantom spectra (a) original spectra and (b) mean-centered spectra.

Some outcomes could be achieved after computing for PCs with PCA as follows. Firstly, PCA results demonstrated that the first four PCs could explain 99.5 % of the variance of the spectral data as 55.44 % for PC1, 30.40 % for PC2, 10.28 % for PC3 and 3.48 % for PC4. Hence, they were chosen to represent the whole data set. Each PC was represented through a loading vector.

Secondly, the loadings of each PCs in Figure 10 can present the contribution of each wavelength (variable) on the original spectral data as explained in Section 4.3. From Figure 10a, the important peaks of PC1 present at the wavelengths 450 nm and 540 nm. Figure 10b shows that PC2 is mainly relating to a peak at 590 nm. Then, the peak at 630 nm contributes significantly to PC3 as in Figure 10c. Finally, Figure 10d presents the major peak for PC4 is about 455 nm. Those peaks are considerably equivalent to the contribution of white, yellow, red, and blue LEDs respectively, when comparing to their single spectra.

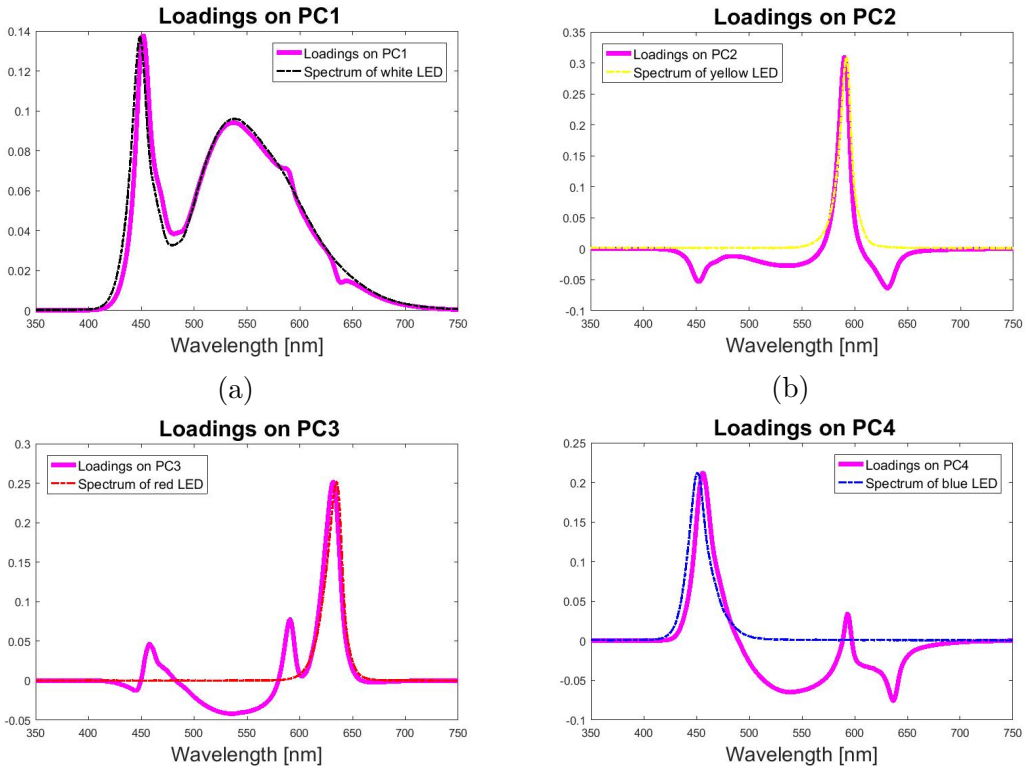


Figure 10: Loadings on the first four PCs from the spectral data of LEDs. (a) Loadings on PC1, (b) Loadings on PC2, (c) Loadings on PC3 and (d) Loadings on PC4.

Thirdly, the variation of the samples is exhibited through the scores on each PC as presented in Figure 11. The score results illustrate the significance of each PC to each sample. For example, from Figure 11c, it can be concluded that the PC3 contributed considerably to samples 31, 34, and 35 compared to other samples.

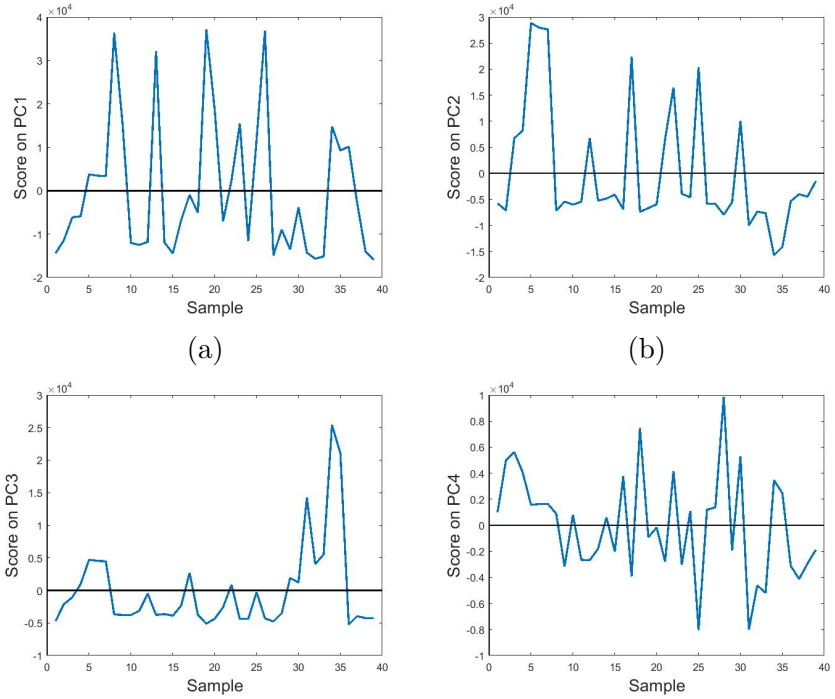


Figure 11: Score plots of 40 LED samples.
(a) Scores on PC1, (b) Scores on PC2, (c) Scores on PC3 and (d) Scores on PC4.

Choosing two samples 4 and 35 as the demonstrations for the computation of PCs. The results are illustrated in Figure A1 (Appendix A). To the left, in the first column, the raw spectra of samples 4 and 35 are presented. The second column shows the mean spectrum (in light blue colour) which is similar to all samples. The third column illustrates mean-centered spectra of the the two samples (in similar colours) with the original spectra. The spectra of the two samples are composed of the PCs. The first PC describes most of the variation in the spectral data. The second PC represents the second most variance in the data set. The third and fourth PCs continue the same manners. Those PCs are prevalent in all samples. After choosing the first fourth PCs, the remained information constructed the residuals which differ among all 40 samples as presented in the last column with cyan colour.

Then, the scores for the first PC versus other chosen PCs were plotted in Figure 12. The information could be induced for two exemplified samples 4 and 35, illustrated as red points in the Figure. Sample 4 included blue, yellow and red LEDs while sample 35 consisted of white and red LEDs. From scores in Figure A1 and score plots in Figures 12a and 12c, sample 4 is contributed mostly by PC1, PC2 and PC4. Figure 12b shows that PC3 plays a minor role in sample 4. It can be interpreted that sample 4 is composed mostly from the white or blue and yellow LEDs and a small amount of red LED. On the other hand, sample 35 is contributed significantly by PC1 as the white LED and PC3 as the red LED. Figure 12b shows that PC3 plays a more significant role in sample 35 compared to sample 4.

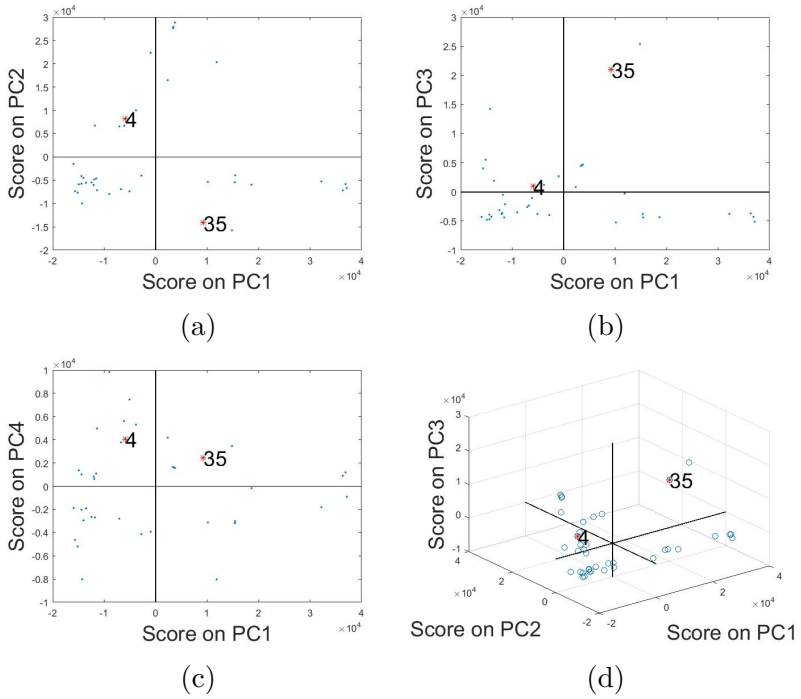


Figure 12: Scores of the PCs plotted against each other for 40 LED samples. (a) Scores on PC1 vs PC2, (b) Scores on PC1 vs PC3, (c) Scores on PC1 vs PC3, and (d) Scatter plot of scores of the first three PCs. Red dots represent the exemplified samples.

5.3 Phantom with colour liquids

In this second phantom, three types of food colour liquids including red, blue and green were dissolved in tap water with known concentrations. Seventeen samples were prepared with only one colour or combinations of colours such as red and blue, blue and green, green and red or all three of them. Each sample was prepared with a different concentration in approximately 10 mL liquid per sample such as 1 % of green colour and 1 % of blue colour, 3 % of red colour with 1 % of green and 1 % of blue. Each colour liquid was transferred into a 55-mm diametre Petri dish (polystyrene) and placed between the light source and the integrating sphere. The system was configured as shown in Figure 13.

5.3.1 Spectral acquisition for colour liquids

Before measurement, the system was warmed up and calibrated as described in Subsection 5.2.1. The integration time was set to approximately 37 ms. Ten scans were averaged to obtain a single spectrum. Averaging scans reduced noise from the spectrum [56]. Boxcar number was set to 1 as the number of adjacent pixels was used to smooth the spectrum with spatial average. Then, a reference spectrum and a dark spectrum were obtained similarly to Figure 7. Prior to each acquisition, those spectra were acquired to minimise the variation of the light source.

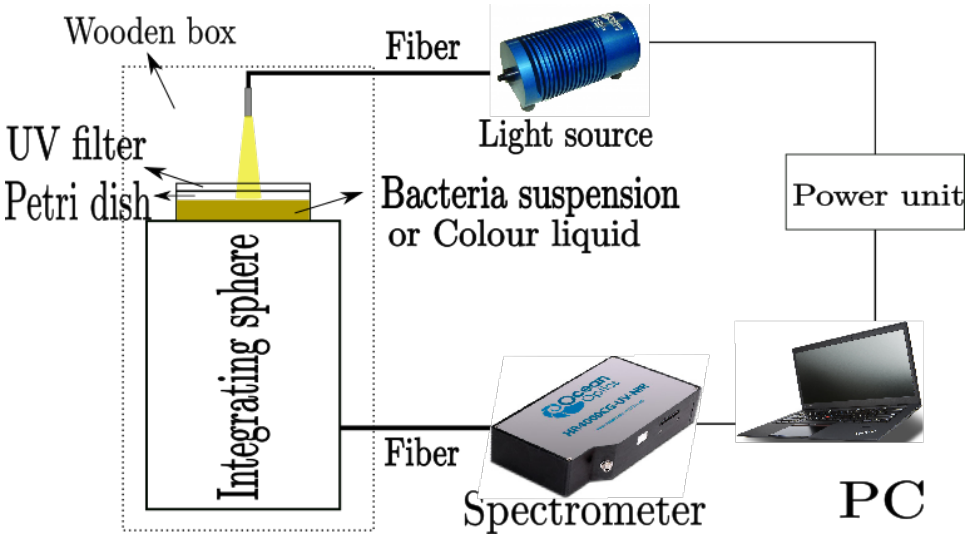


Figure 13: Colour liquid or bacteria measurement system set up.

After the measurement, the absorbance spectra were computed with MATLAB using the following formula [56]:

$$A(\lambda) = -\log_{10} \frac{S(\lambda) - D(\lambda)}{R(\lambda) - D(\lambda)} \tag{22}$$

where

- $A(\lambda)$ is the absorbance at λ [AU],
- $S(\lambda)$ is the measured intensity at λ after the light traversed through the sample,
- $R(\lambda)$ is the intensity of reference spectrum at λ,
- $D(\lambda)$ is the intensity of dark spectrum at λ.

The intensity of reference and dark spectra remained constants in all measurements. Equation (22) demonstrates that when the intensity $S(\lambda)$ decreases, the absorbance $A(\lambda)$ increases and vice versa.

5.3.2 Estimated colour liquid concentration results

The acquired spectra were preprocessed with to retain important information. Firstly, they were down-sampled at approximately 1 nm interval. Secondly, the obtained spectra in this case were affected by the variant of the light source. The beginning and the end of the spectra included mostly noise. Hence, they were truncated so that the remained spectra included only wavelengths from 390 nm to 900 nm. Then, the number of useful wavelengths reduced to 502 measurement points. Some examples of the absorption spectra can be seen in Figure 14. From the Figure, it can be seen that the red colour liquid absorbed most of the light of which wavelengths are below 600 nm. Its spectrum contained three noticeable peaks at 485 nm, 520 nm

and 560 nm. The green colour liquid absorbed most of the wavelengths shorter than 700 nm and included four distinguishable peaks at 405 nm, 450 nm, 480 nm and the strongest one at 630 nm. The blue one absorbed strongly the light at the wavelengths between 400 nm and 420 nm as well as 510 nm and 700 nm with two peaks at 410 nm and 630 nm.

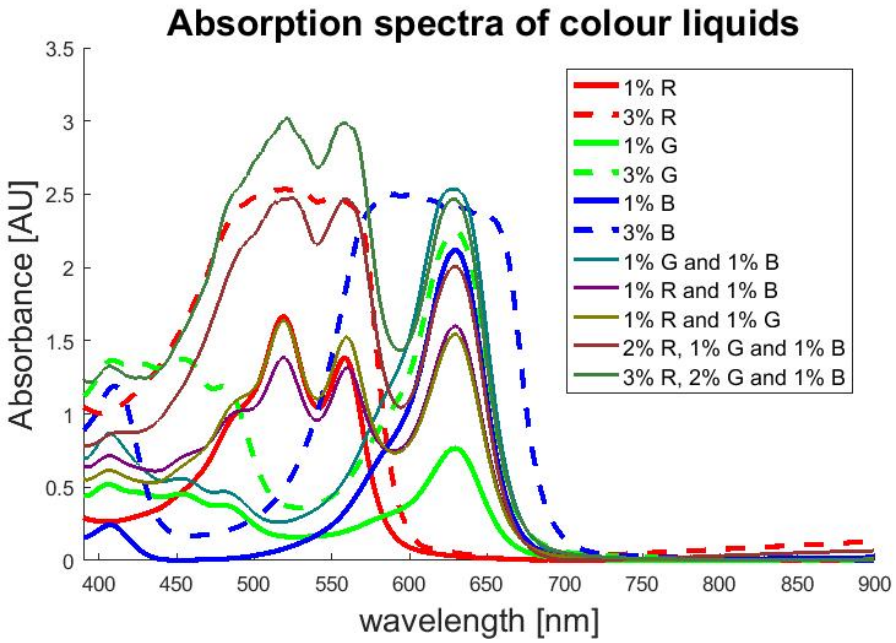
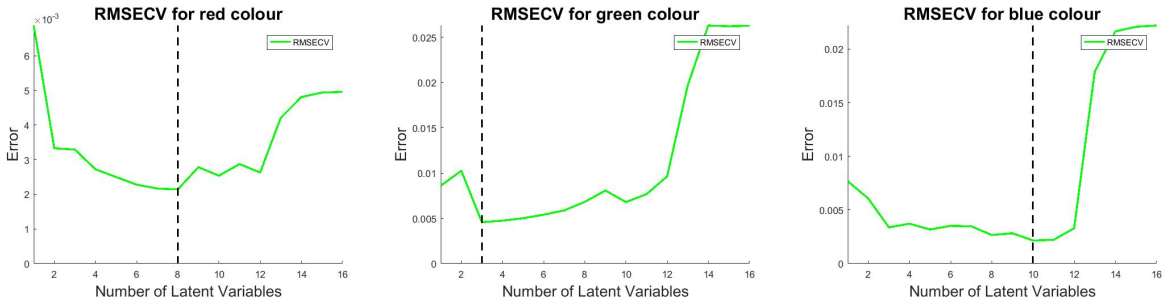


Figure 14: Absorption spectra of colour liquids.
R–Red, G–Green and B–Blue.

After preprocessing, the spectral data was subjected to PLSR in a provided PLS-toolbox [59] for developing the prediction models of colour liquids. The spectra were mean-centered within the computation of the toolbox. Due to the small number of samples, cross validation was utilised to find a suitable model as well as an appropriate amount of latent variables (LVs). The spectra were separated into testing and validating groups using the leave-one-out method since the number of observation was low and it would not prolong the computation time. The RMSEC and RMSECV were calculated to choose an appropriate number of LVs. From the result in Figure 15, it could be concluded that a prediction model including from 4 to 8 LVs was suitable for predicting liquid including red colour. A model with 3 LVs was sufficient for the prediction model for green colour. Another model with between 3 and 10 LVs was appropriate for predicting liquid containing blue colour. Then the model was used for predicting the concentration of each colour in the colour liquids. Figure 16 exemplifies two prediction results with 3 LVs. In this case, the predicted values were almost similar to the true values. However, due to overlapping peaks at 630 nm of both green and blue colour liquid, in many cases, the predicted concentration of those two colour liquids could be interchanged.

The performances of the developed models was validated with the leave-one-out cross validation. The result is illustrated in Table 2. The RMSECV could not be computed because there were liquids with zero concentration of at least one of the colours. It could be concluded that there was no significant improvement when the number of LVs was 8 and 10 for the red and blue colours, respectively; thus, we could choose the lower number of LVs.



(a) (b) (c)

Figure 15: RMSEC and RMSEP to choose LVs for liquid concentration models (a) for red colour, (b) for green colour and (c) for blue colour.

dataYu | dataYuTrue — 2x3 double

	1	2	3
1	0.0291	0	0
2	0.0100	0.0200	0.0100

(a)

dataYu | dataYuTrue — 2x3 double

	1	2	3
1	0.0258	0.0051	0.0011
2	0.0082	0.0230	0.0096

(b)

Figure 16: Prediction results of liquid concentrations (a) true concentrations and (b) predicted concentrations of two samples. Column 1 – blue colour, column 2 – green colour and column 3 – red colour.

Table 2: The performances of prediction models for colour liquid concentrations.

Model	**Colour**	**Variable**	**LVs**	**Calibration**		**Cross-validation**	
				R_C^2	RMSEC	R_{CV}^2	RMSECV
PLSR	Red	502	8	0.993	0.0011	/	0.0022
	Green	502	3	0.930	0.0037	/	0.0046
	Blue	502	10	0.981	0.0019	/	0.0022
PLSR	Red	502	4	0.980	0.0018	/	0.0027
	Green	502	3	0.930	0.0037	/	0.0046
	Blue	502	3	0.944	0.0033	/	0.0034

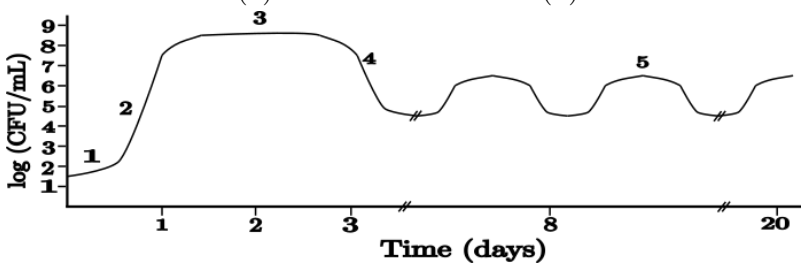
6 Bacteria experiment implementation

This chapter describes the experiment on bacterial suspension, the main target of this project. The set up system described in Section 5 was applied to measure the bacteria with some necessary modifications. The chapter describes the procedures of culturing bacteria, preparing the suspension, modifying the measurement system, obtaining spectra and analysing spectral data with the concentrations of bacteria.

6.1 Bacterial culture

Escherichia coli (*E. coli*) strain K-12 HB101 (Bio-Rad) for education was used in the project. This specific strain is well-adapted to laboratory environment and suitable for studying and researching purposes, because it is generally recognised as safe. *E. coli* (Figure 17a) was chosen as the model organism for this project because it is common in the lower intestine of warm-blood organisms like humans and it has been extensively studied and well understood in biological research [60]. This type of bacteria is easy to grow with basic nutrient and environmental requirements [60]. Most of its growth cycle (Figure 17c) including lag, exponential (log), and stationary phases frequently occurs within 24 hours in optimal growing conditions at 37 °C. This property allows data acquisition in short time, and experiment repetition can be easily executed. This project mainly focused on the exponential phase since it is frequently studied in other research. The death phase of *E. coli* follows after several days, and it was only used for a minor analysis in this project.

(a) (b)



(c)

Figure 17: *E. coli* (a) morphology [61], (b) colonies on a Petri dish and (c) typical shape of its growth curve [62].
1–lag phase, 2–log phase, 3–stationary phase,
4–death phase, and 5–long-term stationary phase.

Bacterial culturing procedure was conducted through the following steps:

1. Luria-Bertani (LB) broth and LB broth plates with 2 % agar were prepared beforehand with LB powder including 10 g triptone, 10 g NaCl and 5 g yeast extract in 1 L of Milli-Q water.
2. A small amount of bacteria, stored at −70 °C, was transferred from a 2 mL cryovial, onto a Petri dish containing sterile LB broth agar.
3. The bacteria on the plate were incubated upside down overnight (Incubator LabRum Klimat, Germany, Figure 18) at 37 °C to produce a starting plate. From this plate (Figure 17b), individual bacterial colonies were taken for further experiments. Each colony is a cluster of microorganism cloned from a single cell. Preparing the starting plate allowed sufficient time for the bacteria to rehydrate from a frozen condition and adapt to the culture medium. This significantly reduced time for the lag phase or equivalently waiting time of the experiments when the bacteria were transfered into culturing tubes later.
4. Ten well-isolated colonies were incubated in LB broth culture medium for 24 hours rotating at 140 rpm and 37 °C. This step granted an adequate population of bacteria and assured that bacteria were distributed evenly in the medium.
5. After 24 hours, bacteria aliquot was transferred into 200 mL LB broth to start the culturing and measurement process in the log phase within 6 hours. This growth phase was targeted since it is usually applied in biological studies.
6. Bacterial suspension was maintained for 15, 24, and 72 hours for stationary phases measurement.
7. The death phase spectra were acquired after 7, 8 and 20 days while the suspension was maintained rotating at room temperature. In this case, the temperature did not affect the growth in this phase of the bacteria.

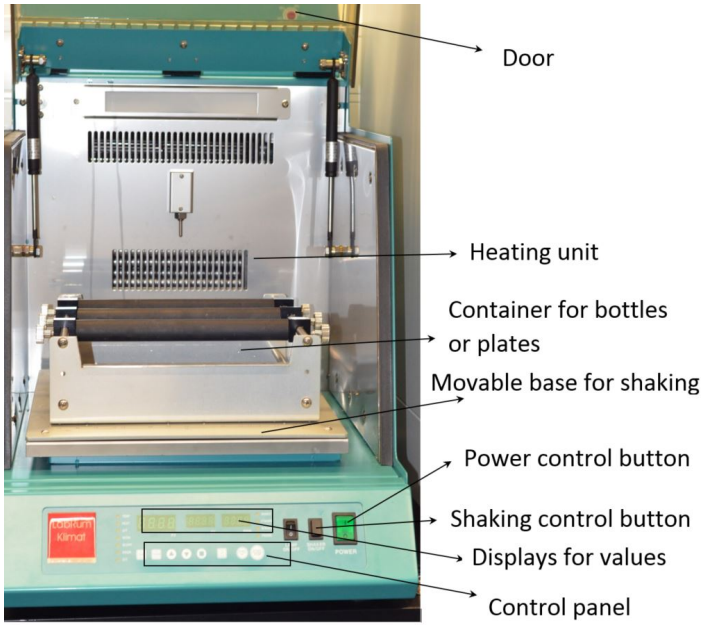


Figure 18: LabRum Klima incubator.

6.2 Referencing

In this project, the spectral acquisition was a secondary method, meaning that it could not directly measure the bacterial concentration but required a referencing method. Two methods, optical density and serial dilution, were utilised as the references since they are typically used for cell estimating and counting in microbiological standard.

Optical density

Optical density (OD) measures the relative absorbance of a liquid sample of an analyte in comparison to a blank sample. Here, OD was measured against pure LB broth as the blank sample. Each sample of 2 mL of bacterial suspension was transferred into a cuvette (VWR semi-micro) with a standard optical path length of 10 mm using a pipette. It was measured at 600 nm with a VWR UV 1600PC single-beam spectrometer (range 190–1100 nm) as shown in Figure 19 [63]. The spectrometer was set to absorbance measurement mode. This method was chosen as a reference since it can prove the growth of bacteria in the lag and exponential phases. The method can also classify the stationary phase and the death phase of bacteria. When bacteria reach the stationary phase, OD number remains stable over a long period of time. The death phase happens when the value of OD is reduced compared to the previous stage. During the experiment, the variation of the OD value could be checked by placing the cuvette into the holder and taking it out repeatedly. An example could be seen from Figure 20. Several relative observations can be made:

- OD < 0.2, the value was almost stable.
- $0.2 <$ OD < 0.6, it varied ± 0.02 units between when it was placed into the holder and when it became stable.
- Between 0.6 and 0.9, the OD value remained stable.
- OD > 0.9, the value reduced after a few minutes since the bacteria sank to the bottom of the cuvette. Hence, the even distribution could not be guaranteed and might result in some significant errors.
- In the stationary and death phases, OD was approximately 1.5 ± 0.1 units.

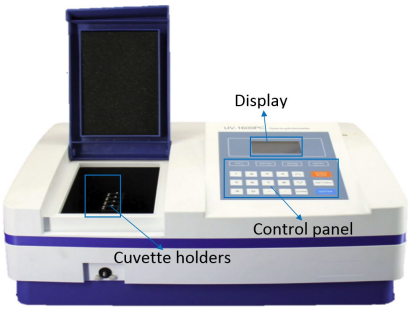


Figure 19: VWR UV 1600PC single-beam spectrometer.

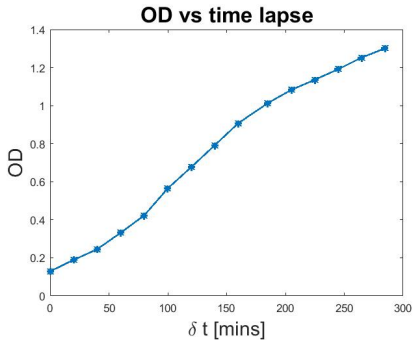


Figure 20: OD vs Time.

Serial dilution

Serial dilution is another standard method to manually quantify bacteria concentration in microbiology. Its working principle is analogous to OD. This method can confirm the growth phases of bacteria through the number of colony forming units per millilitre (CFU mL^{-1}). When bacteria are in the log phase, the number of CFU increases over time. During the stationary phase, the value stands at a certain level. Finally, the value drops in the death stage. Additionally, it is capable of detecting contaminated samples.

Serial dilution can be illustrated as in Figure 21. In the project, 100 µL sample aliquots were 10-fold serially diluted in saline (0.9 % NaCl liquid). Then 100 µL was spread with an L-stick on a Petri dish (90-mm diametre) and incubated 24 hours. Each sample was triplicated to statistically reduce the errors. A plate with saline was used as the control. Then, the plates were manually counted for colony forming units (CFU). The most statistically valid plates were the ones with the CFU between 20 and 200 and they were chosen to compute original bacterial concentration [60]. The number of formed colonies were recorded as CFU mL^{-1}. The following formula was applied for calculating CFU [60]:

$$CFU = \frac{\text{number of colonies} \cdot \text{dilution factor}}{\text{plated volume}} \tag{23}$$

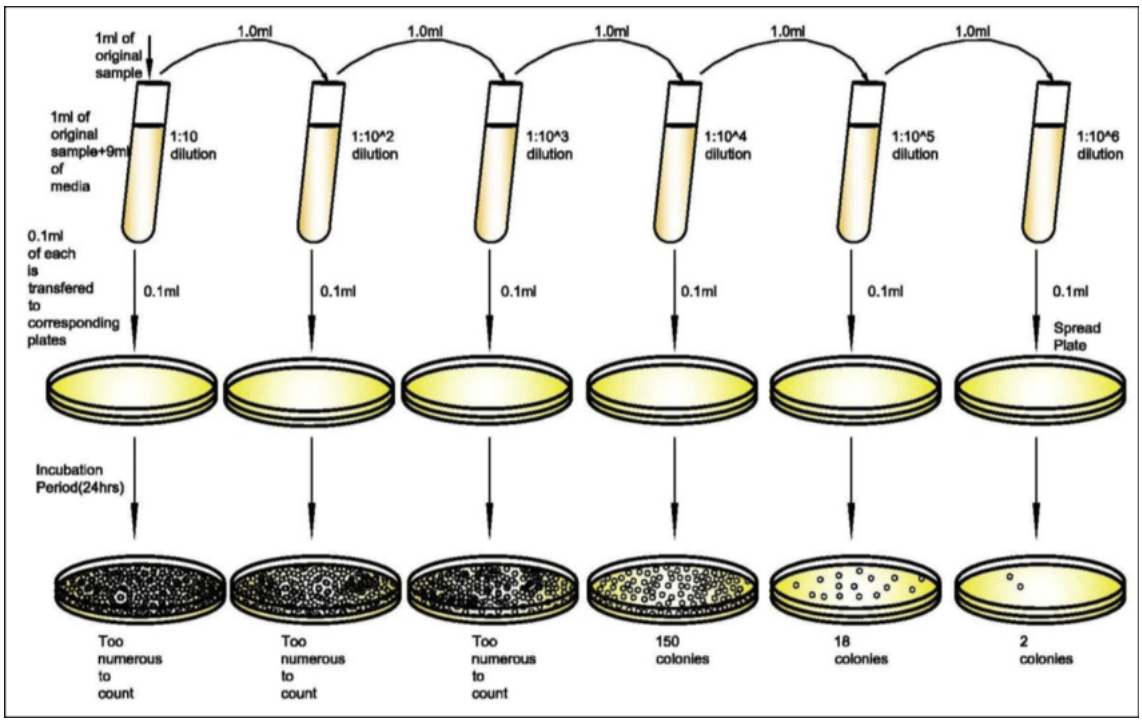


Figure 21: Serial dilution process [64].

In the following cases, the counted results were anomalous and discarded:

- One of the triplicated dishes from the same fold included a significantly higher or lower CFU than the others. It could be the result of taking samples unevenly.
- The number of counted CFU was significantly higher for an OD value within a series of increasing OD values.
- Similar OD numbers corresponded to obviously different values of CFU mL^{-1}. The abnormal value was stem from the manual error of laboratory process.

6.3 Spectral acquisition

All experiments were performed at room temperature, approximately 24 °C. Before implementing measurements, a calibration process was performed with the halogen light source and blank space between the light source and the integrating sphere. The integration time was selected manually between 2400 ms and 3000 ms. The reference spectrum from the halogen source and the dark spectrum were captured and stored similar to the implementation in Subsection 5.2.1.

Some other apparatuses were also employed to provide an optimal measurement environment. Firstly, an UV-filter which limited the intensity of wavelengths less than 390 nm was placed above the Petri dish during each measurement. It reduced the killing effect of UV light on the bacteria. Secondly, the entire measurement system was located rigidly inside a wooden box with a lid as can be seen in Figure 22. The inner side of the box and the bottom of the lid were covered with black clothes or painted with black color to prevent the reflection from its walls. Its outer side was wrapped with aluminum fold to block the disturbance of other ambient light. The system was set up as in Figure 13. The upper fibre for transmitting light was located at the closest possible position to the integrating sphere, approximately 22 mm. A Petri dish and the UV-filter could be laid between them without touching. This set up allowed to capture the most intensive light beam from the source.

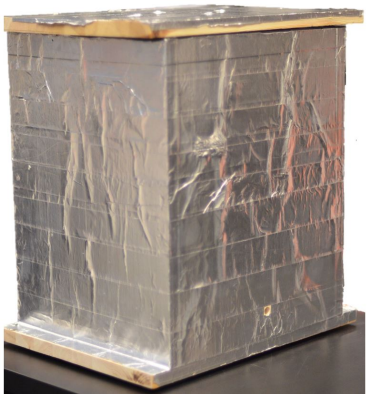


Figure 22: Wooden box covered with aluminum fold to block ambient light.

After overnight culturing, 10 mL bacterial suspension was transferred into a bottle including 200 mL of new LB broth. Then it was incubated for about 30 minutes while shaking at 140 rpm and 37 °C so that the temperature of the liquid could stabilise and bacteria started to reproduce. Each measurement was taken at 30 minutes time intervals. In each measurement, 12 mL bacterial suspension was transferred from the 200 mL vial to a 55-mm diametre Petri dish with a pipette. This particular Petri dish was chosen because it could lie stably on the top of the integrating sphere. The suspension samples were transferred into the dish inside a laminar flow hood. Flowing air could maintain the sterility of the suspension since it moved bacteria and dust in the ambient atmosphere away from the experimental zone. Then the bacterial suspension was cooled down and prevented from the optimal condition for bacterial replication. Additionally, cooling down also helped to avoid the evaporation of the liquid which could have blurred the lid of the Petri dish and possibly reduced total light beam traversing into the measurement system.

In order to facilitate the spectral acquisition, a simple interface was established using trigger function of SpectraSuite. The interface encompassed five buttons: blank, dark, Petri, LBliquid and bacteria as shown in Figure 23. They were equivalent to reference spectra, dark spectra, spectra of Petri dish, spectra of LB liquid and spectra of bacteria. When pressing each button, a spectrum was saved to a pre-specified location.

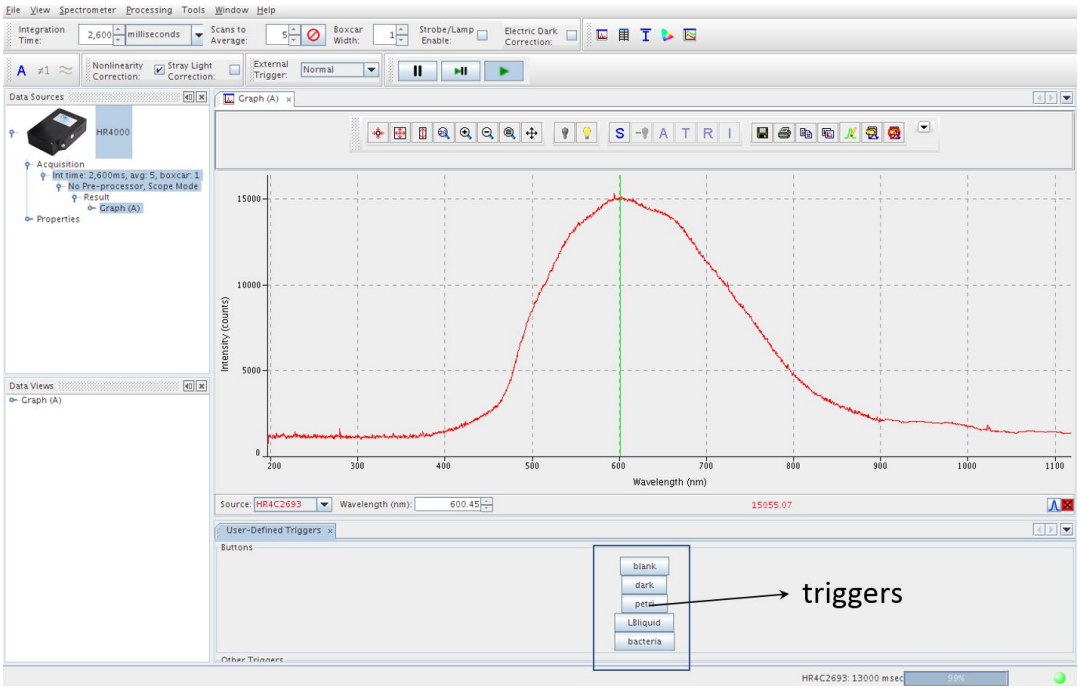
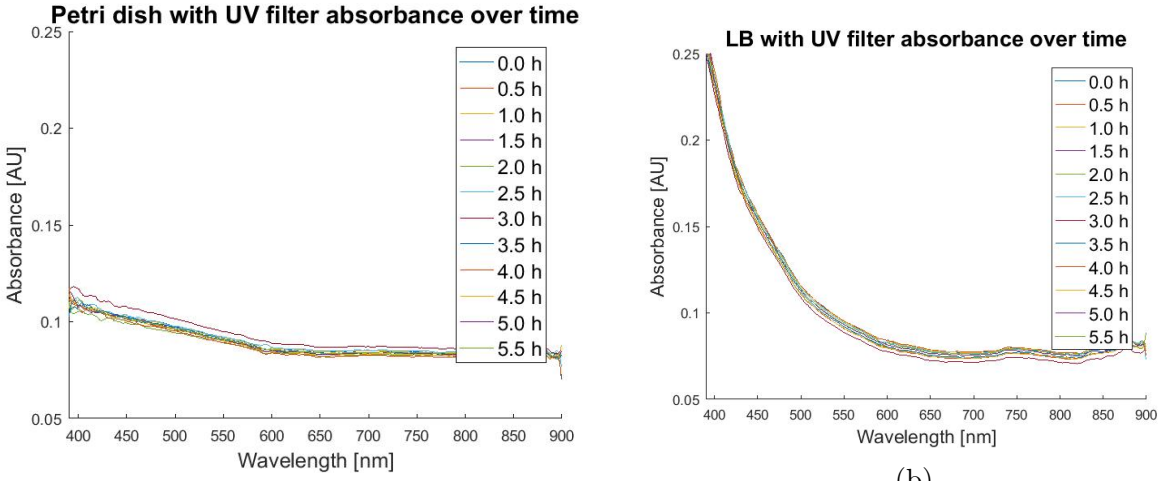


Figure 23: SpectraSuite interface with triggers to save spectral files.

Before starting the experiment, the optical system was set up and recalibrated as described in Section 5.2. Each measurement started with measuring the light source spectrum and dark spectrum since the light source slightly changed over time. This

step maintained the consistency of the system. Then, an empty Petri dish was measured just before taking the sample. Next, a bacterial sample was filled into the Petri dish. After a waiting time of five minutes, it was placed between the light beam and the integrating sphere. For each sample, five different positions on the dish were chosen randomly to measure the spectra. Two consecutive spectra were captured for each position. Capturing spectra at different positions proved that the bacteria distributed evenly in the suspension and the spatial arrangement did not strongly affect the spectral analysis. A mean of all ten spectra would be calculated and inputted as a single spectrum into the analysis procedure. After that, another Petri dish containing pure LB liquid was also measured. The spectra of Petri dishes were acquired throughout the entire experiment for comparing purpose. Similar spectra of Petri dish and LB liquid verified that the existence of bacteria in suspension resulted in spectra difference in comparison to the uncontaminated liquid. Their typical spectra are illustrated in Figure 24a and 24b, respectively. Besides, the absorbance spectrum of Milli-Q water was also obtained to compare the analysis results in the end of the project.



(a) (b)

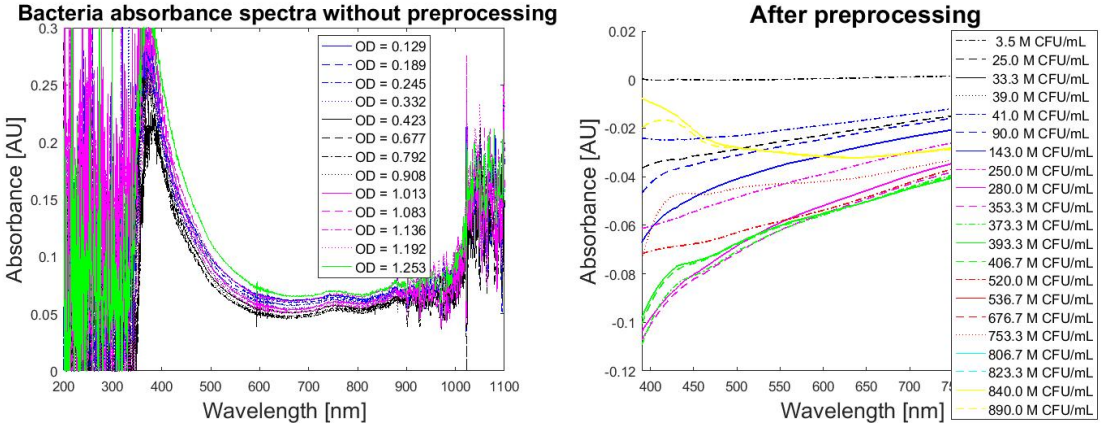
Figure 24: (a) Absorbance spectra over time of Petri dish with UV filter and (b) Absorbance spectra over time of LB liquid in Petri dish with UV filter. Integration time = 2400 ms.

6.4 Spectral preprocessing

After spectral acquisition process, the absorbance spectra were computed with equation (22) using MATLAB and subjected to some preprocessing methods. The acquired spectra were down-sampled as described in Subsection 5.2.2 and truncated to remove noise at two ends of the spectra similarly as Subsection 5.3.2. Then each spectrum consisted of 502 variables.

From those spectra, no distinguishable peak could be found. Due to high acquisition time, a peak at the wavelength of 595 nm appeared to be noise. It was considered as noise because the peak also occurred in the spectra of empty Petri dish and LB liquid or it did not appear consistently in all spectra of the bacteria. Hence, the peak needed to be removed because it could lead to some errors in later calculation, and a smoothing method was applied for this purpose. It was executed with function smooth of MATLAB using a window length of 71.

Besides, the absorbance spectra of the Petri dish and LB liquid varied slightly within a day of the experiments. Their variation was more noticeable between different days of the experiments due to the variations of light source and culture medium. Hence, the absorbance spectra of LB liquid were subtracted from the absorbance spectra of bacteria in order to reduce the variation. Finally, the whole spectral data was subjected to PCA and PLS for computing PCs and modelling, respectively.



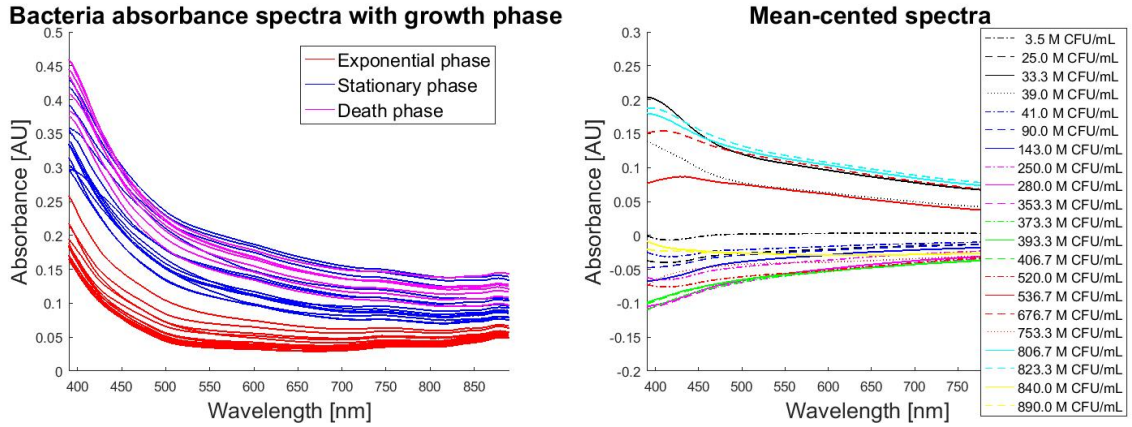
(a) (b)

Figure 25: Absorbance spectra of *E. coli* (a) before and (b) after preprocessing.

6.5 PCA for spectral clustering

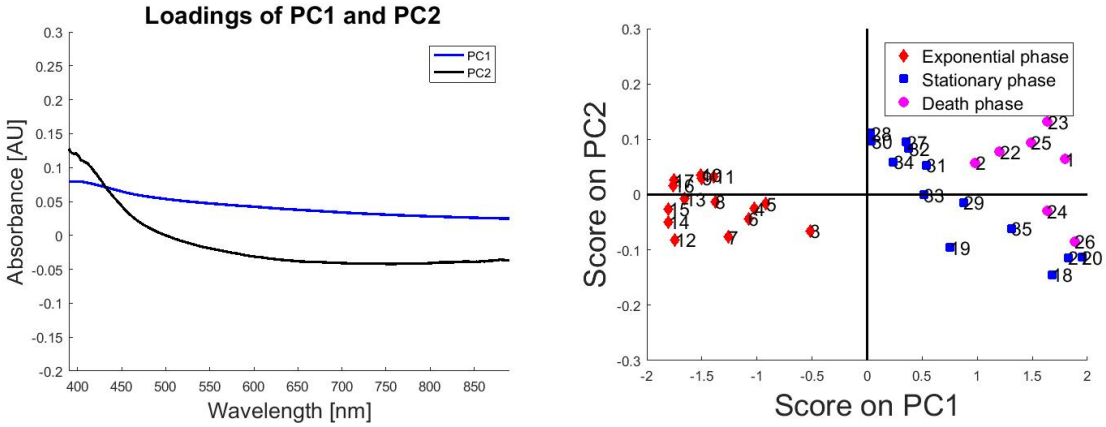
PCA was adopted to reduce the dimensionality of the spectral data and to determine the possibility of identifying the growth phases of *E. coli*. The acquired spectra of bacteria in different growth phases (Figure 26a) were subjected to pca function of MATLAB for computing PCs. The mean was subtracted from the whole data set within the function execution. Mean-centered spectra are presented in Figure 26b.

After computing with PCA, the analysis results demonstrated that loadings of PC1 and PC2 accounted for 99.87 % of the total variation (99.32 % and 0.54 %, respectively). PCs captured the variance in the data set and clustered the spectra into groups with presumptively similar spectral properties. Score plots in Figure 27b show clustered results from the spectra of *E. coli* in the exponential, stationary and death phases.



(a) (b)

Figure 26: (a) Absorbance pectra of *E. coli* at different growth phases and (b) Mean-centered spectra.



(a) (b)

Figure 27: (a) PC1 and PC2 loadings for absorbance spectra of *E. coli* and (b) PCA for spectra of *E. coli* at different growth phases.

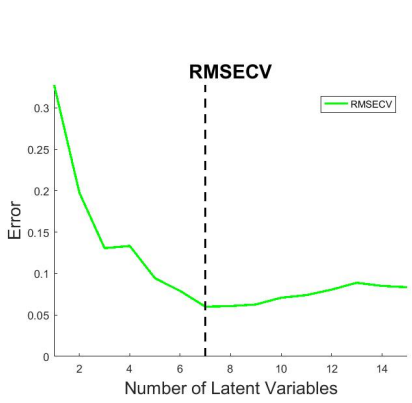
6.6 PLS for quantitative model

PLS was used according to a presumption that during the exponential phase, the number of CFU grew almost linearly with time. Additionally, it has been applied in bacterial quantification as mentioned in Section 4.4. PLS was applied to investigate two types of associations. The first one was between the spectra obtained from the spectrophotometer and OD from the second spectrophotometer. The second association was between the spectra and the total viable count (TVC).

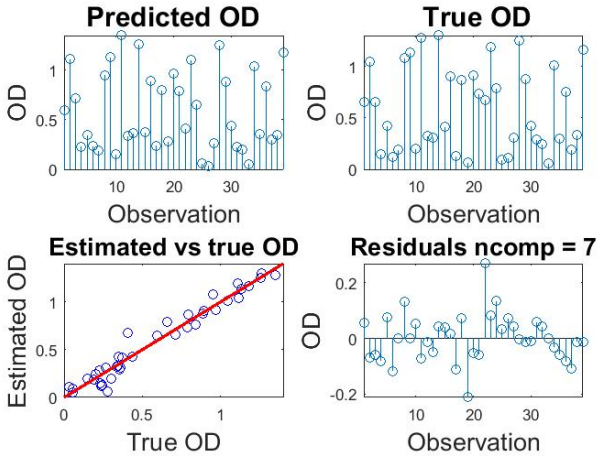
The preprocessed spectra were subjected to pls function for modelling. Due to a low number of samples, leave-one-out cross validation was performed to evaluate the model. This was done by removing one sample (test sample) from the whole data set and the PLRS model was established for the remaining samples (training samples). The whole data set was randomised so that any dominant trend existing in the spectra could not critically influence the prediction results. The minimum of

RMSECV was utilised to find the number of PLS latent variables (LVs). Then, the model was evaluated based on the determination coefficient R^2 and the standard error of calibration (RMSEC) and validation (RMSECV).

The prediction results are illustrated in Figure 28 for the OD and Figure 29 for the CFU. For the OD prediction model, the number of LVs was seven in accordance with the lowest RMSECV in Figure 28a. For the CFU prediction model, six LVs were selected as illustrated in Figure 29a.

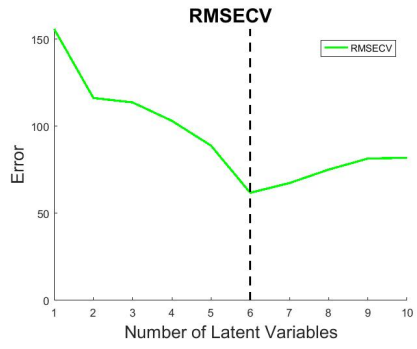


(a)

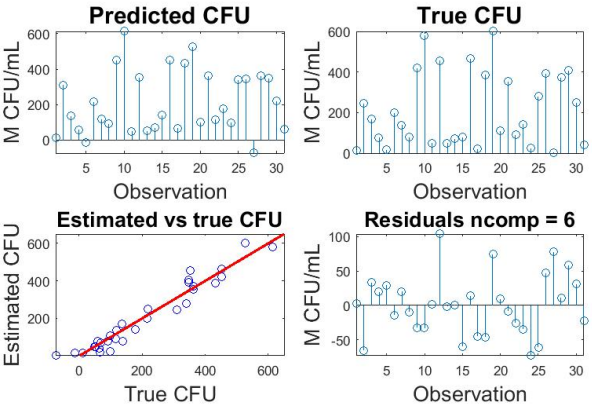


(b)

Figure 28: Predicted OD (a) RMSECV for choosing LVs and (b) Predicted results.



(a)



(b)

Figure 29: Predicted TVC (a)RMSECV for choosing LVs and (b) Predicted results.

6.7 Optimal wavelengths selection

The preprocessed spectral data was standardised by dividing it with the standard deviation of each wavelength as illustrated in Figure 30. The same analytic procedure was conducted similar the description in Section 6.6. Other prediction steps for standardised data could be seen in Appendix B. Then, the weighted regression coefficients were plotted to find the critical wavelengths with function findpeaks of MATLAB and manually. The new model was reevaluated with leave-one-out cross validation.

Figure 31 and Figure 32 illustrate the important wavelengths which can be used for predicting the OD and the CFU, respectively. For each OD prediction, nine wavelengths were chosen at 463, 607, 532, 641, 651, 680, 834, 857 and 870 nm. For CFU prediction, eight wavelengths with the highest weights were chosen at 398, 423, 462, 520, 602, 861, 872 and 881 nm. The result will be discussed in Section 7.3 to compare between the model established with full spectra and this model.

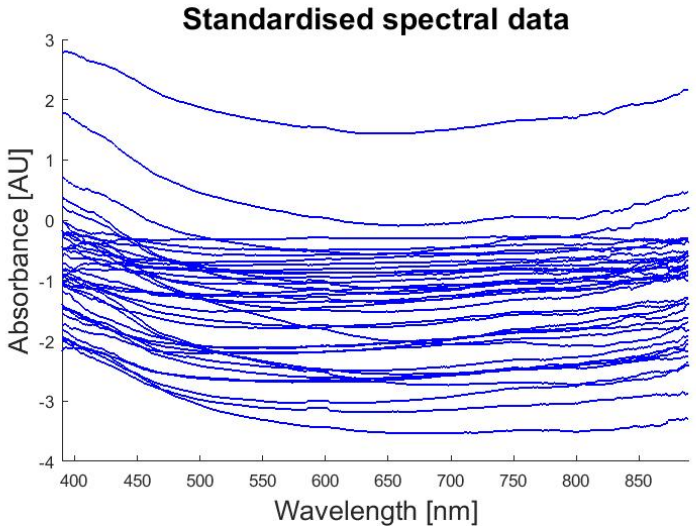


Figure 30: Standardised absorbance spectra of *E. coli*.

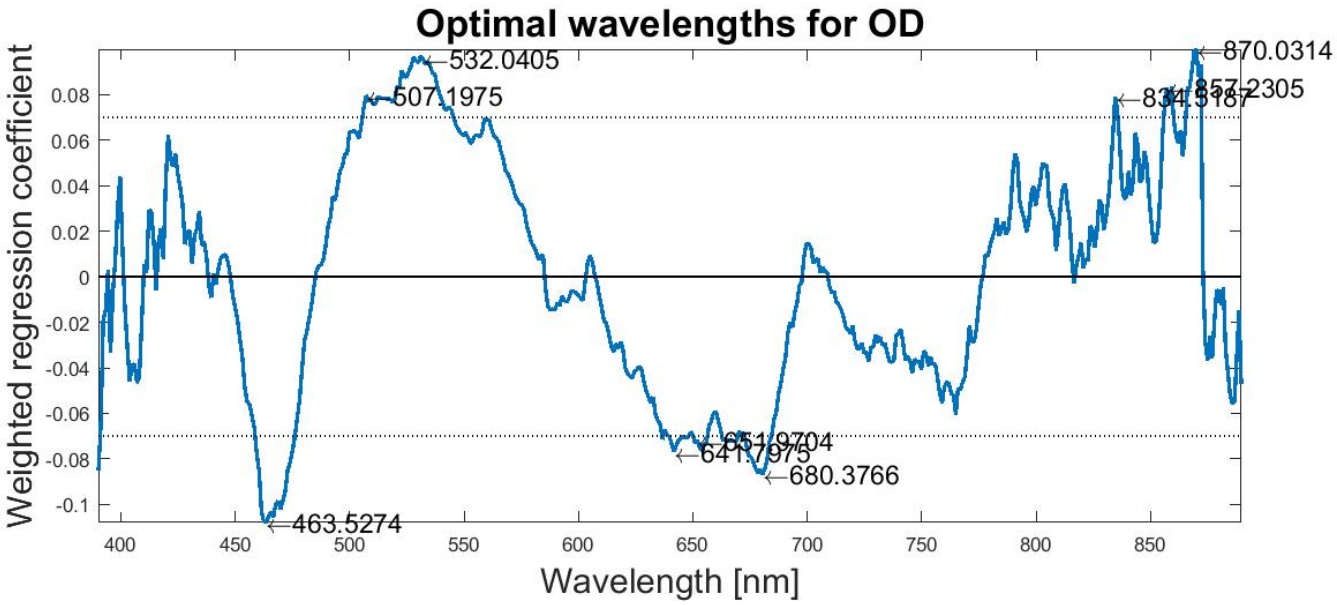


Figure 31: Weighted regression coefficients and optimal wavelengths for OD prediction.

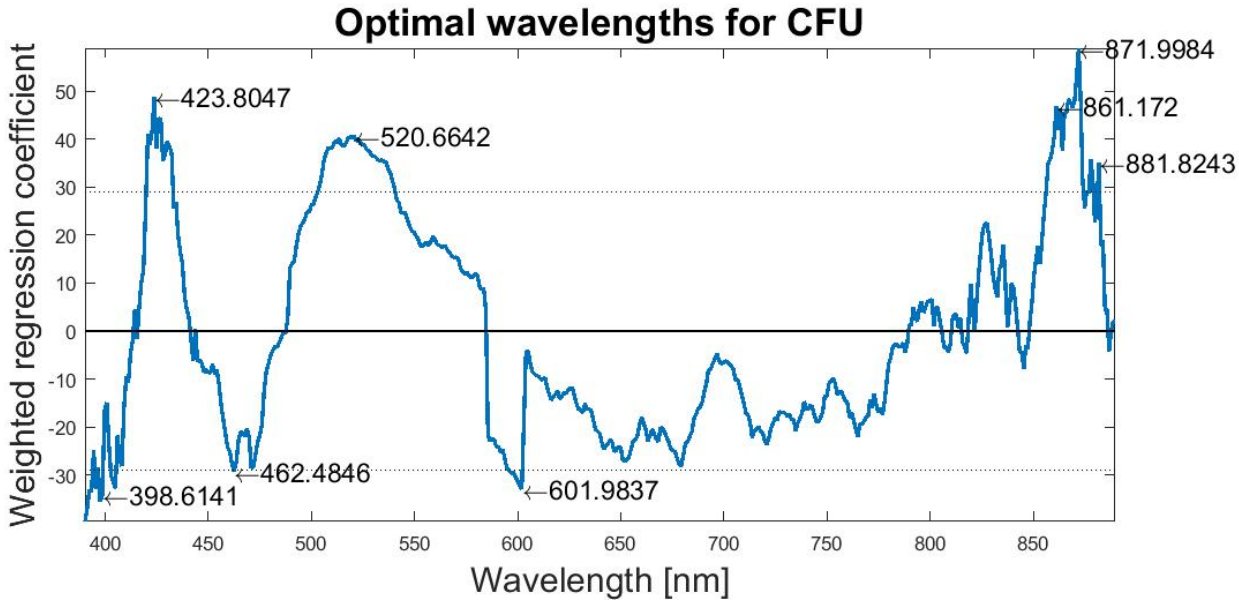


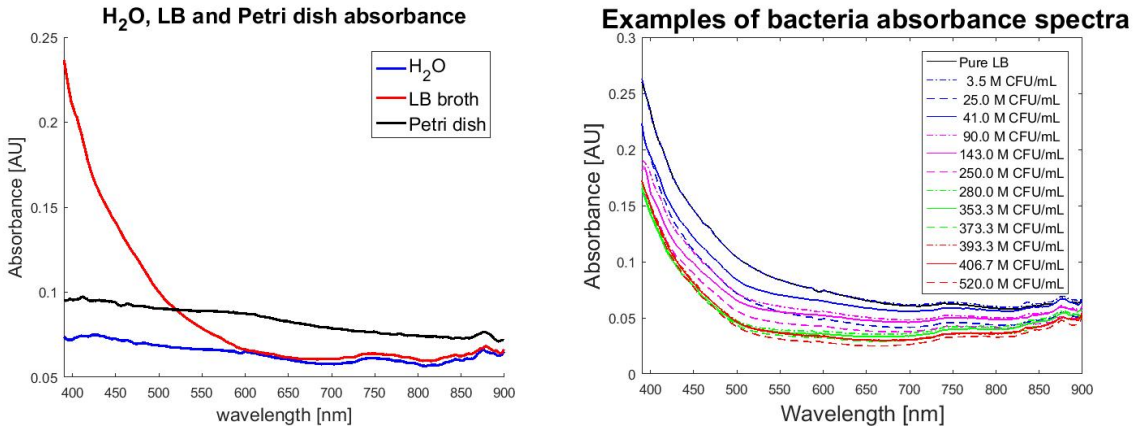
Figure 32: Weighted regression coefficients and optimal wavelengths for CFU prediction.

7 Discussion on bacterial analysis results

7.1 Spectral shape and trend

The bacteria were monitored when growing in the culturing medium and no complicated sample preparation was required in the project. From the acquired spectra of Petri dish and LB liquid over time in Figure 24, it could be concluded that the variation of the bacterial spectra was due to the existence of cells in the liquid.

The absorbance spectrum of Luria-Bertani broth medium and pure water almost overlaps in the range from 640 nm to longer wavelengths as can be seen from Figure 33a. It can also be seen that the culturing medium attenuates significantly the low end of the spectrum [24]. This spectral shape result is consistent with the result from the previous studies [24][65]. This attenuating property can strongly influence the characteristic peaks of bacteria if they exist.



(a) (b)

Figure 33: (a)Absorbance spectra of water, LB broth and Petri dish with UV filter (b) Distinguishable absorbance spectra of bacteria with UV filter.

The obtained spectra illustrated that there was no absorption peak existing in the UV-VIS-NIR region for *E. coli*. This could be concluded by comparing them with the absorption spectra of colour liquids, which demonstrated some particular absorption peaks in this range. This result was in agreement with the previous studies by Alupoaei and Garcia-Rubio (2004) [65], Kiefer *et al.* (2010) [24], and Myers *et al.* (2013) [25]. Hence, any wavelengths in this range can be used to measure the absorbance.

During the log phase, the absorption of light reduced gradually. It may be a result of material consumption of bacteria in the medium. When the OD was between 0.8 and 1.0, the linear manner could not be guaranteed since the corresponding spectra almost overlapped each other.

From Figure 33b, no significant difference could be found between the spectra of

bacteria with concentration about 4×10^6 CFU mL^{-1} and pure Luria-Bertani liquid. The sytem, thus, could not detect bacteria when the concentration was lower than this limit. Hence, the sensitivity was not high in comparison to other studies where the concentration level could be as low as 10^5 CFU mL^{-1} such as in the study from Alupoaei (2004) *et al.* [26] and Kiefer *et al.* (2010) [24] using a similar spectral range, or 10^3 CFU mL^{-1} in the research of Lu (2011) *et al.* [47] and Nakakimura *et al.* (2012) [12] using mid-infrared range.

7.2 Growth phases clustering

Clustering the spectra of different growth phases was implemented to verify whether the spectroscopic method could differentiate between the spectra since the project targeted only on the exponential phase of bacteria. The transients between those phases were not encompassed in this research since they can complicate the model. The spectra were visually and analytically distinct between the exponential, stationary and death phase of the bacteria as shown in figures 26a and 27b. Literally, the lag phase was included in the exponential phase since it lasted for a short period. The analytical result could not distinguish the lag phase from the exponential phase. The spectra of the stationary and death phases were almost similar since the byproducts do not disappear when the bacteria die, and materials were released from the broken cells as explained in Section 3.2. This result was in line with the result measured from the VWR spectrometer when the optical density also demonstrated a minor fluctuation.

The cluster result is illustrated in Figure 27b. It can be noticed that the first PC played a dominant role in the whole spectral data. Hence, the exponential phase can be significantly distinguished from the stationary and death phases. The two latter phases might not be classified from each other. They were more likely to be contributed by the second PC when comparing the Loadings of PC2 with their spectral shapes in Figure 27a. However, the minor role of this PC was insufficient to considerably discriminate them.

7.3 Quantitative results

The predicted OD results in Figure 28 illustrate that the model was efficient in predicting OD between 0.2 and 0.9. Within this range, the OD number increased quickly over a few minutes and fluctuated ± 0.02 units from the experimental observations. When the OD was less than 0.1, the predicted values were often higher than the real value. It could be explained as the bacterial suspension was in lag phase and did not vary much from the spectrum of the pure medium. When the OD was above 0.9, the bacteria were in the end of the exponential phase or in the stationary phase, the OD

number decreased slightly over time because the biomass of the bacteria was high and cells easily sank to the bottom of the cuvette as well as the Petri dish. Hence, the spectra at this range could result in errors of the model.

Table 3 presents the predicted result for OD. It can be seen that the general range of the errors was up to 0.07 units which was higher than the variation range of the OD from VWR spectrometer (± 0.02). However, the R^2 values showed that the model had a high prediction capability. The optimum model, WRC-PLSR, achieved comparable results with the PLSR model. The optimum model could slightly reduce the number of LVs but increase the RMSE. Hence, it could reduce the computational effort compared to a model constructed from a full scan, but it could result in more errors.

Table 3: The performances of prediction model for OD.

Model	**Variable**	**LVs**	**Calibration**		**Cross-validation**	
			R^2_C	RMSEC	R^2_{CV}	RMSECV
PLSR	502	7	0.992	0.064	0.853	0.067
WRC-PLSR	9	6	0.986	0.077	/	0.068

The predicted concentration CFU mL^{-1} is presented in Figure 29. Theoretically, absorbance based on light scattering indirectly measures CFU but it is closely related to the dry weight of cells [25]. The quantitative results were also affected by the variations of light sources. However, according to Myers *et al.*(2013) [25], as long as the same spectrometer is used, it is possible to determine the concentration of bacteria within some standard deviations. The precision of the model was dependent on the results of the laboratory work. Some reference values of the plate count were discarded as explained in Section 6.2.

When an optimum model for predicting CFU was constructed with the optimal wavelengths, its errors slightly increased but did not influence its prediction capability. For the WRC-PLSR model, the number of LVs reduced by one in comparison with the PLSR model, as illustrated in Table 4. Therein, the WRS-PLSR model can be used with some acceptable deviations when the computational resource is limited.

Table 4: The performances of prediction models for CFU.

Model	**Variable**	**LVs**	**Calibration**		**Cross-validation**	
			R^2_C	RMSEC	R^2_{CV}	RMSECV
PLSR	502	6	0.980	40.08	/	61.75
WRC-PLSR	8	5	0.948	65.36	/	81.61

8 Conclusions and future work

This thesis presents an empirical project on modelling the growth of bacteria for the application of photodynamic therapy with indocyanine green and near-infrared light. The main target was establishing a model for quantifying the concentration of bacteria in a liquid sample. The project has been capable of developing a model which correlated the spectral response with the concentration of bacterial cells.

In order to establish the model, the research applied a spectroscopic method to estimate the concentration of bacteria in the culturing medium. The approach was based on light absorption and scattering by the particles in a liquid sample. Collected spectral data was analysed using PCA and PLS to extract the meaningful information. Principal component analysis (PCA) was utilised to reduce the dimensionality of the data and to cluster the spectra of *Escherichia coli* strain K-12 for different growth phases into relevant groups. Partial least square (PLS) was applied to establish a model relating the spectra to the concentration of bacteria in the culturing medium. This spectroscopy approach is one of the most popular methods used in estimating the concentration of analytes in chemistry and biology. Hence it was an acceptable approach for this study. The spectra in the study were limited to 390–900 nm because the study aimed to develop a basic system which could determine the concentration of bacteria and the system can be applied in a future work to measure the intensity of near-infrared light at the wavelengths 805–810 nm. However, this range limits the feasibility to distinguish dead cells from live cells.

At the beginning of the project, two phantom experiments were performed using LED lights and colour liquids in order to verify the feasibility of the spectroscopic approach. The phantoms yielded evidences demonstrating that the spectroscopic approach could be employed under controlled conditions. Then, the bacterial suspensions were placed into the system. The spectra of bacteria were acquired over time. They were correlated with the concentrations of bacteria using PLS. The true concentrations were determined with traditional methods including optical density and total viable count. These two methods were utilised as the references for quantification since they are normally the most standard methods in microbiology.

Some significant results could be concluded after conducting experiments. Firstly, the spectra proved that there were no characteristic peaks for *E. coli* within the spectral range 390–900 nm, which was in agreement with other studies [24] [25] [65]. Hence, any wavelengths in this range can be chosen for absorbance measurement. Secondly, when the concentration of bacteria was higher than 4×10^6 CFU mL^{-1}, this spectroscopy method could quickly detect their existence in the culture medium within a short period of time. The results suggested that UV-VIS-NIR spectroscopy can be used in detecting bacteria in liquid. However, the sensitivity of this system was not too high in comparison with other studies [12] [47]. Thirdly, the results

achieved from PCA could be used as input for classifying the growth phases of *E. coli* using the first two components; thus, the dimensionality of the spectral data could be significantly reduced. Finally, the obtained spectra could be used for developing a prediction model for concentration of *E. coli* in the exponential phase with the upper limit of about 6×10^8 CFU mL^{-1}.

The study was only conducted for estimating the concentration of bacteria and limited to *E. coli* as the target. Hence the project should be repeated using other types of bacteria for a wider range of utility. Additionally, a longer spectral range may be utilised for a better quantification of bacteria. In comparison to other studies and methods, this project used a larger amount of bacterial suspension for each measurement than the frequently used cuvette; thus, it produced more waste. However, the wider area of Petri dish compared to the cuvette could be useful for PDT since it could capture more beam from the light source and be easier to apply indocyanine green. For further applicability, it would be beneficial to determine the association between the concentration of indocyanine green (ICG) and the power of near-infrared (NIR) light in photodynamic therapy (PDT) with the viability of bacteria.

References

[1] C. L. Ventola, "The antibiotic resistance crisis: part 1: causes and threats," *Pharmacy and Therapeutics*, vol. 40, no. 4, p. 277, 2015.

[2] I. A. Rather, B.-C. Kim, V. K. Bajpai, and Y.-H. Park, "Self-medication and antibiotic resistance: Crisis, current challenges, and prevention," *Saudi journal of biological sciences*, vol. 24, no. 4, pp. 808–812, 2017.

[3] E. Martens and A. L. Demain, "The antibiotic resistance crisis, with a focus on the united states," *The Journal of antibiotics*, vol. 70, no. 5, p. 520, 2017.

[4] N. Topaloglu, M. Güney, S. Yuksel, and M. Gülsoy, "Antibacterial photodynamic therapy with 808-nm laser and indocyanine green on abrasion wound models," *Journal of biomedical optics*, vol. 20, no. 2, 2015.

[5] T. J. Dougherty, C. J. Gomer, B. W. Henderson, G. Jori, D. Kessel, M. Korbelik, J. Moan, and Q. Peng, "Photodynamic therapy," *JNCI: Journal of the National Cancer Institute*, vol. 90, no. 12, pp. 889–905, 1998.

[6] J. T. Alander, I. Kaartinen, A. Laakso, T. Pätilä, T. Spillmann, V. V. Tuchin, M. Venermo, and P. Välisuo, "A review of indocyanine green fluorescent imaging in surgery," *Journal of Biomedical Imaging*, vol. 2012, p. 7, 2012.

[7] S. Prahl. Optical absorption of indocyanine green (ICG). [Accessed 26.04.2017]. [Online]. Available: http://omlc.ogi.edu/spectra/icg/index.html

[8] N. Chiniforush, M. Pourhajibagher, S. Parker, S. Shahabi, and A. Bahador, "The *in vitro* effect of antimicrobial photodynamic therapy with indocyanine green on *Enterococcus faecalis*: Influence of a washing vs non-washing procedure," *Photodiagnosis and photodynamic therapy*, vol. 16, pp. 119–123, 2016.

[9] F. Vatansever, W. C. de Melo, P. Avci, D. Vecchio, M. Sadasivam, A. Gupta, R. Chandran, M. Karimi, N. A. Parizotto, R. Yin, G. P. Tegos, and M. R. Hamblin, "Antimicrobial strategies centered around reactive oxygen species–bactericidal antibiotics, photodynamic therapy, and beyond," *FEMS microbiology reviews*, vol. 37, no. 6, pp. 955–989, 2013.

[10] *Metrohm*, "NIR spectroscopy. A guide to near-infrared spectroscopic analysis of industrial manufacturing processes," White Paper, *Metrohm NIR system*, 2013.

[11] C. Quintelas, D. P. Mesquita, J. A. Lopes, E. C. Ferreira, and C. Sousa, "Near-infrared spectroscopy for the detection and quantification of bacterial contaminations in pharmaceutical products," *International journal of pharmaceutics*, vol. 492, no. 1, pp. 199–206, 2015.

[12] Y. Nakakimura, M. Vassileva, T. Stoyanchev, K. Nakai, R. Osawa, J. Kawano, and R. Tsenkova, "Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations," *Analytical Methods*, vol. 4, no. 5, pp. 1389–1394, 2012.

[13] T. Kiesslich, A. Gollmer, T. Maisch, M. Berneburg, and K. Plaetzer, "A comprehensive tutorial on *in vitro* characterization of new photosensitizers for photodynamic antitumor therapy and photodynamic inactivation of microorganisms," *BioMed research International*, vol. 2013, April 2013.

[14] L. Beytollahi, M. Pourhajibagher, N. Chiniforush, R. Ghorbanzadeh, R. Raoofian, B. Pourakbari, and A. Bahador, "The efficacy of photodynamic and photothermal therapy on biofilm formation of *Streptococcus mutans*: An *in vitro* study," *Photodiagnosis and Photodynamic therapy*, vol. 17, pp. 56–60, March 2017.

[15] E. A. Genina, A. N. Bashkatov, G. V. Simonenko, O. D. Odoevskaya, V. V. Tuchin, and G. B. Altshuler, "Low-intensity indocyanine-green laser phototherapy of acne vulgaris: pilot study," *Journal of Biomedical Optics*, vol. 9, no. 4, pp. 828–834, 2004.

[16] A.-M. Mamoon, A. M. Gamal-Eldeen, M. E. Ruppel, R. J. Smith, T. Tsang, and L. M. Miller, "*In vitro* efficiency and mechanistic role of indocyanine green as photodynamic therapy agent for human melanoma," *Photodiagnosis and Photodynamic Therapy*, vol. 6, no. 2, pp. 105–116, 2009.

[17] W. Bäumler, C. Abels, S. Karrer, T. Weiss, H. Messmann, M. Landthaler, and R. Szeimies, "Photo-oxidative killing of human colonic cancer cells using indocyanine green and infrared light," *British Journal of Cancer*, vol. 80, no. 3-4, pp. 360–363, 1999.

[18] O. Bozkulak, R. F. Yamaci, O. Tabakoglu, and M. Gulsoy, "Photo-toxic effects of 809-nm diode laser and indocyanine green on MDA-MB231 breast cancer cells," *Photodiagnosis and Photodynamic therapy*, vol. 6, no. 2, pp. 117–121, 2009.

[19] K. Krumova and G. Cosa, "Chapter 1 Overview of reactive oxygen species," in *Singlet oxygen: applications in Biosciences and Nanosciences, Volume 1.* The Royal Society of Chemistry, 2016, vol. 1, pp. 1–21.

[20] P. Held, "An introduction to reactive oxygen species. Measurement of ROS in cells," White Paper, Jan 2014.

[21] T. A. Dahl, W. Midden, and P. E. Hartman, "Comparison of killing of gram-negative and gram-positive bacteria by pure singlet oxygen." *Journal of Bacteriology*, vol. 171, no. 4, pp. 2188–2194, 1989.

[22] (2013) *ICNIRP guidelines in limits of exposure to incoherent visible and infrared radiation.* International Commission On Non-ionizing Radiation Protection. [Online]. Available: http://www.icnirp.org/cms/upload/publications/ICNIRPVisible_Infrared2013.pdf

[23] P. S. Yarmolenko, E. J. Moon, C. Landon, A. Manzoor, D. W. Hochman, B. L. Viglianti, and M. W. Dewhirst, "Thresholds for thermal damage to normal

tissues: an update," *International Journal of Hyperthermia*, vol. 27, no. 4, pp. 320–343, 2011.

[24] J. Kiefer, N. Ebel, E. Schlücker, and A. Leipertz, "Characterization of escherichia coli suspensions using uv/vis/nir absorption spectroscopy," *Analytical Methods*, vol. 2, no. 2, pp. 123–128, 2010.

[25] J. A. Myers, B. S. Curtis, and W. R. Curtis, "Improving accuracy of cell and chromophore concentration measurements using optical density," *BMC biophysics*, vol. 6, no. 1, p. 4, 2013.

[26] C. E. Alupoaei, J. A. Olivares, and L. H. Garcia-Rubio, "Quantitative spectroscopy analysis of prokaryotic cells: vegetative cells and spores," *Biosensors and Bioelectronics*, vol. 19, no. 8, pp. 893–903, 2004.

[27] D.-W. Sun, *Modern techniques for food authentication*. Academic Press, 2008.

[28] H. P. R. Aenugu, D. S. Kumar, N. P. Srisudharson, S. S. Ghosh, and D. Banji, "Near infra-red spectroscopy–an overview," *International Journal of ChemTech Research*, vol. 3, no. 2, pp. 825–836, 2011.

[29] J. T. Alander, V. Bochko, B. Martinkauppi, S. Saranwong, and T. Mantere, "A review of optical nondestructive visual and near-infrared methods for food quality and safety," *International Journal of Spectroscopy*, vol. 2013, 2013.

[30] H. C. Hulst and H. C. van de Hulst, *Light scattering by small particles*. Courier Corporation, 1957.

[31] M. Kerker, *The scattering of light and other electromagnetic radiation*. Elsevier, 2016.

[32] P. J. Wyatt, "Differential light scattering: a physical method for identifying living bacterial cells," *Applied optics*, vol. 7, no. 10, pp. 1879–1896, 1968.

[33] M. Kotlarchyk, S.-H. Chen, and S. Asano, "Accuracy of RGD approximation for computing light scattering properties of diffusing and motile bacteria," *Applied optics*, vol. 18, no. 14, pp. 2470–2479, 1979.

[34] K. Shimizu and A. Ishimaru, "Scattering pattern analysis of bacteria," *Optical Engineering*, vol. 17, no. 2, p. 172129, 1978.

[35] P. Chỳlek and J. Li, "Light scattering by small particles in an intermediate region," *Optics communications*, vol. 117, no. 5, pp. 389–394, 1995.

[36] M. T. Madigan and J. M. Martinko, *Biology of Microorganisms*, 11th ed. NJ: Pearson, 2006.

[37] D. L. Massart, B. G. Vandeginste, L. Buydens, P. Lewi, J. Smeyers-Verbeke, and S. d. Jong, *Handbook of chemometrics and qualimetrics: Part A*, 3rd ed. Amsterdam: Elsevier Science Inc., 2003.

[38] N. Kumar, A. Bansal, G. Sarma, and R. K. Rawal, "Chemometrics tools used in analytical chemistry: An overview," *Talanta*, vol. 123, pp. 186–199, 2014.

[39] T. Dearing. Fundamentals of Chemometrics and Modeling. University of Washington. [Accessed 15.06.2017]. [Online]. Available: http://depts.washington.edu/cpac/Activities/Meetings/documents/DearingFundamentalsofChemometrics.pdf

[40] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.

[41] W. Gander and J. Hrebicek, *Solving problems in scientific computing using Maple and Matlab®*. Springer Science & Business Media, 2011.

[42] L. I. Smith, "A tutorial on principal components analysis," Cornell University, USA, Tech. Rep., February 26 2002. [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[43] I. A. Cowe and J. W. McNicol, "The use of principal components in the analysis of near-infrared spectra," *Applied spectroscopy*, vol. 39, no. 2, pp. 257–266, 1985.

[44] J. Shlens, "A tutorial on principal component analysis," *Cornell University*, 2014.

[45] H. M. Al-Qadiri, N. I. Al-Alami, M. Lin, M. Al-Holy, A. G. Cavinato, and B. A. Rasco, "Studying of the bacterial growth phases using fourier transform infrared spectroscopy and multivariate analysis," *Journal of Rapid Methods and Automation in Microbiology*, vol. 16, no. 1, pp. 73–89, 2008.

[46] N. Nicolaou, Y. Xu, and R. Goodacre, "Fourier transform infrared and raman spectroscopies for the rapid detection, enumeration, and growth interaction of the bacteria *Staphylococcus aureus* and *Lactococcus lactis* ssp. *cremoris* in milk," *Analytical chemistry*, vol. 83, no. 14, pp. 5681–5687, 2011.

[47] X. Lu, H. M. Al-Qadiri, M. Lin, and B. A. Rasco, "Application of mid-infrared and raman spectroscopy to the study of bacteria," *Food and Bioprocess Technology*, vol. 4, no. 6, pp. 919–935, 2011.

[48] F. Cámara-Martos, G. Zurera-Cosano, R. Moreno-Rojas, R. M. García-Gimeno, and F. Pérez-Rodríguez, "Identification and quantification of lactic acid bacteria in a water-based matrix with near-infrared spectroscopy and multivariate regression modeling," *Food Analytical Methods*, vol. 5, no. 1, pp. 19–28, 2012.

[49] M. Kamruzzaman, G. ElMasry, D.-W. Sun, and P. Allen, "Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis," *Analytica Chimica Acta*, vol. 714, pp. 57–67, 2012.

[50] J. P. Wold, T. Jakobsen, and L. Krane, "Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy," *Journal of Food Science*, vol. 61, no. 1, pp. 74–77, 1996.

[51] J. Schneider, "Cross validation," *A Locally Weighted Learning Tutorial Using Vizier*, vol. 1, 1997.

[52] H. J. Seltman, *Experimental design and analysis*. Carnegie Mellon University, 2012.

[53] A. G. Frenich, D. Jouan-Rimbaud, D. Massart, S. Kuttatharmmakul, M. M. Galera, and J. M. Vidal, "Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares," *Analyst*, vol. 120, no. 12, pp. 2787–2792, 1995.

[54] S. D. Osborne, R. Künnemeyer, and R. B. Jordan, "Method of wavelength selection for partial least squares," *Analyst*, vol. 122, no. 12, pp. 1531–1537, 1997.

[55] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A review of variable selection methods in partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 62–69, 2012.

[56] *HR4000 and HR4000CG-UV-NIR series spectrometer installation and operation manual*, Ocean Optics, 2016. [Online]. Available: https://oceanoptics.com/wp-content/uploads/hr4000.pdf

[57] *Halogen light source with attenuator and TTL-shutter installation and operation manual*, Ocean Optics, 2015. [Online]. Available: https://oceanoptics.com/wp-content/uploads/hl2000fhsa1.pdf

[58] I. Robinson, "importoceanoptics," 2015, the MathWorks, Natick, MA, USA.

[59] P. Geladi, "Workbook: Regression for calibration in OCTAVE or MATLAB," 2010.

[60] J. K. Brown, *Biotechnology. A laboratory skills course*, 1st ed. BioRad, 2011.

[61] Biocote. (2016, July) [Accessed 29.09.2017]. [Online]. Available: https://www.biocote.com/blog/five-facts-e-coli/

[62] P. Pletnev, I. Osterman, P. Sergiev, A. Bogdanov, and O. Dontsova, "Survival guide: *Escherichia coli* in the stationary phase," *Acta Naturae*, vol. 7, no. 4 (27), 2015.

[63] *Spectrophotometer UV 1600PC UV-VIS*, VWR, 2013. [Online]. Available: https://us.vwr.com/assetsvc/asset/en_US/id/16641873/contents

[64] K. Neupane, "Bacterial inhibition in waste-water/fracking water using copper ion solution," Ph.D. dissertation, Youngstown State University, 2016.

[65] C. E. Alupoaei and L. H. García-Rubio, "Growth behavior of microorganisms using uv-vis spectroscopy: *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 86, no. 2, pp. 163–167, 2004.

A PCA for LED samples

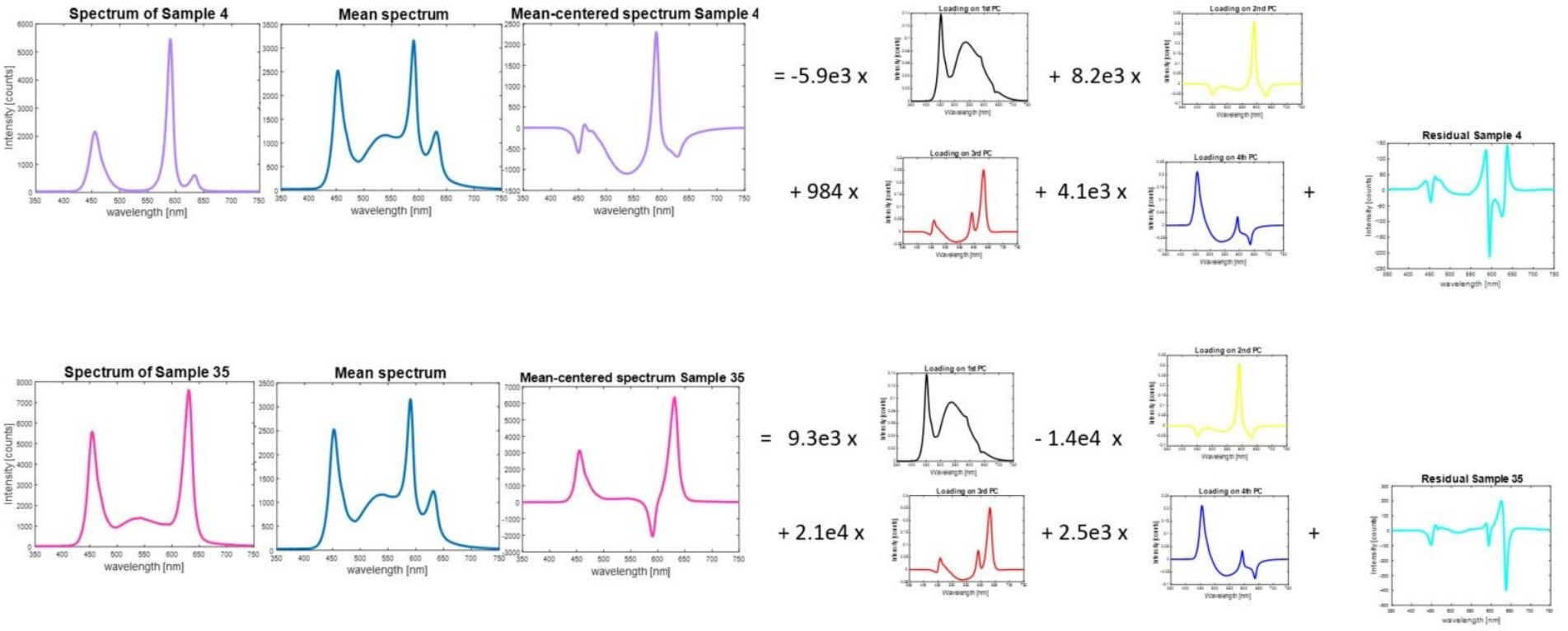


Figure A1: PCA works on two samples 4 (blue, yellow and red) and 35 (white and red) (—) PC1, (—) PC2, (—) PC3, (—) PC4, and (—) residuals.

B Predicted result with optimal wavelengths

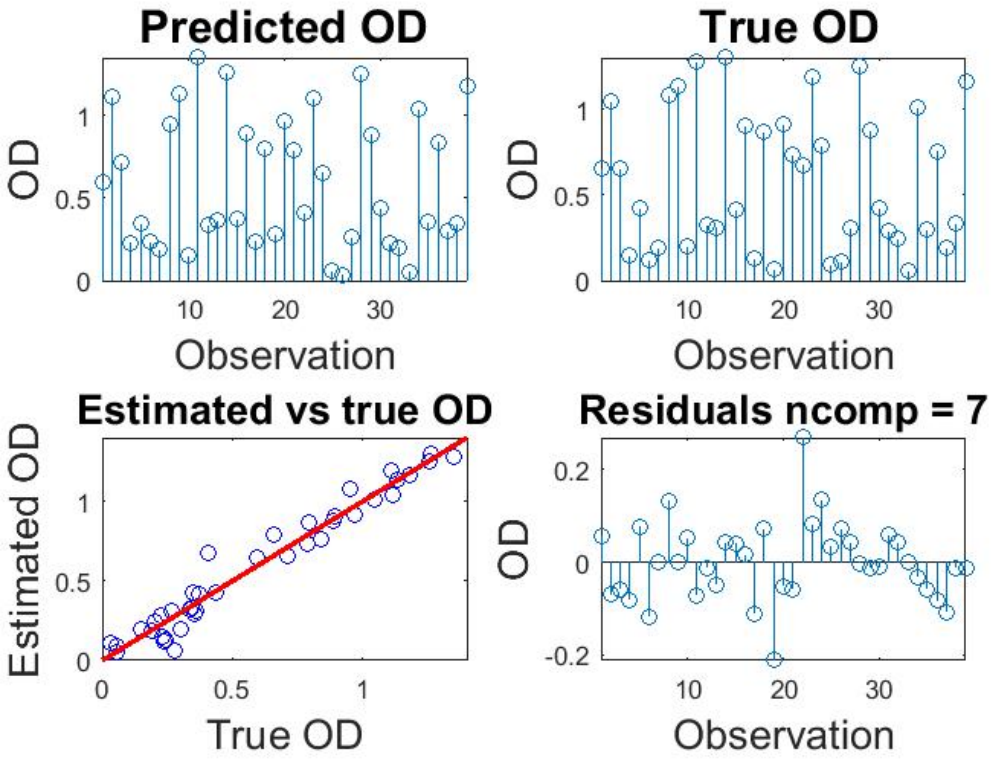


Figure B1: Predicted OD from the original data.

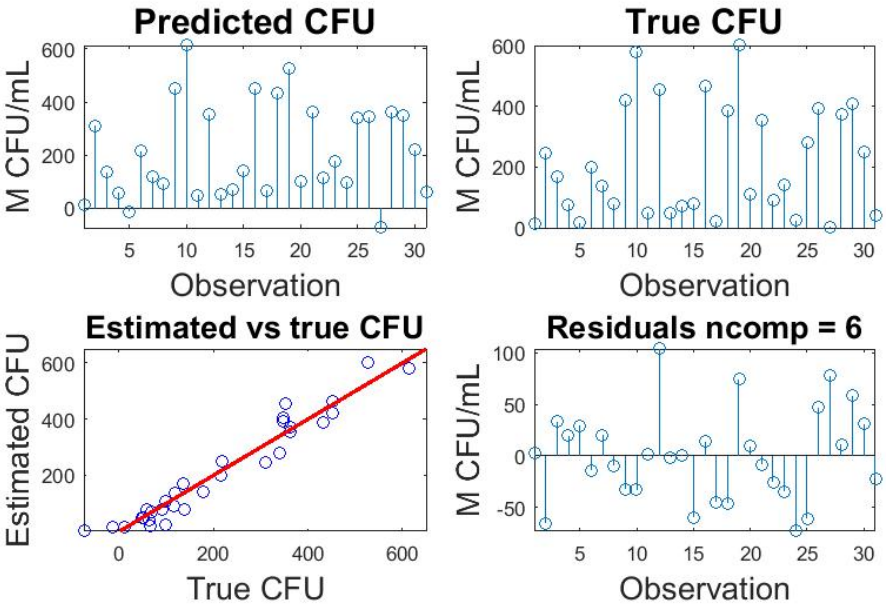
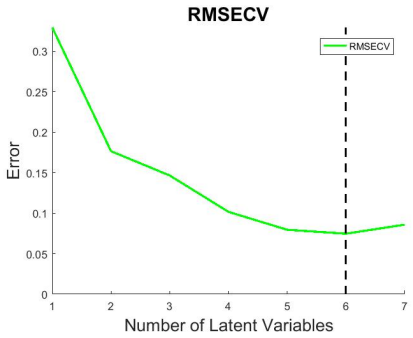
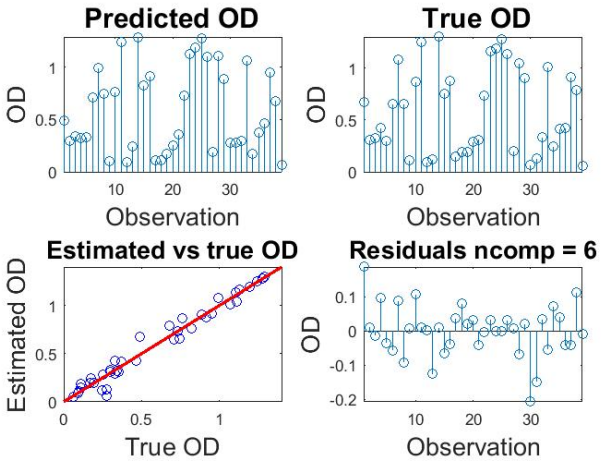


Figure B2: Predicted CFU from the original data.

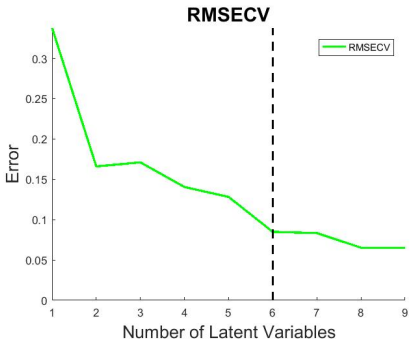


(a)

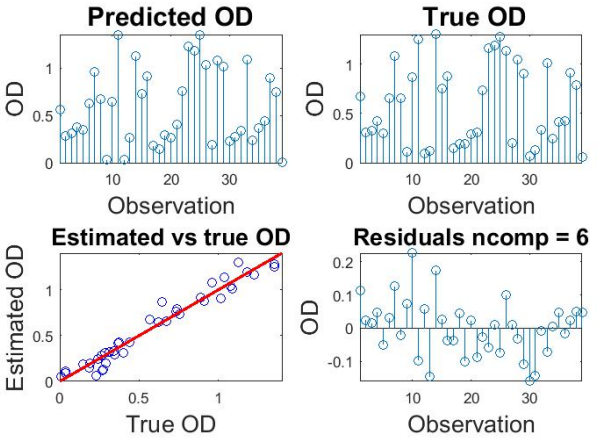


(b)

Figure B3: Predicted OD from the standardised data
(a) RMSECV for choosing LVs and (b) Predicted results.

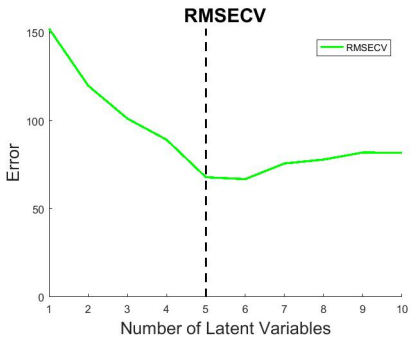


(a)

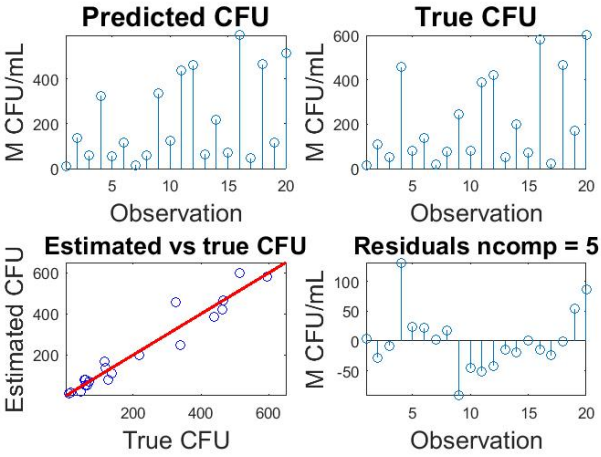


(b)

Figure B4: Predicted OD from the optimal wavelengths
(a) RMSECV for choosing LVs and (b) Predicted results.

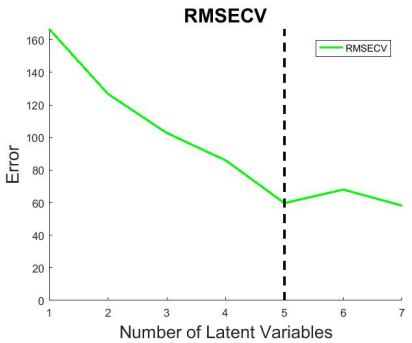


(a)

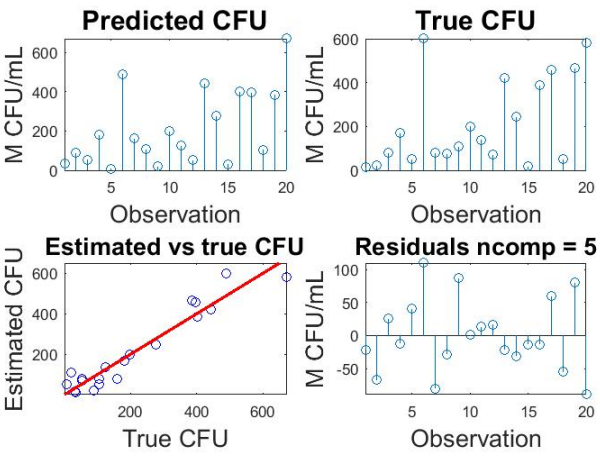


(b)

Figure B5: Predicted CFU from the standardised data
(a) RMSECV for choosing LVs and (b) Predicted results.



(a)



(b)

Figure B6: Predicted CFU from the optimal wavelengths
(a) RMSECV for choosing LVs and (b) Predicted results.

C Total viable counts

The following tables report the Total viable count of *E. coli* from different days. Some notations are used to present the results:

– For the plates included less than 20 or more than 200 colonies or the dilution factor was not plated.
x For the plates were excluded due to some reasons, for example, multiple colonies formed clumps and could not be distinct, blank plate, or part of the plate was dry and the results could not be assured.
Rep. is an abbreviation for Repetition.
Dil. is an abbreviation for Dilution.

Table C1: Total viable count log phase day 1.

Time	10:25	10:45	11:05	11:25	11:45	12:05	12:25	12:45	13:05	13:25
OD	0.129	0.189	0.245	0.332	0.423	0.565	0.677	0.792	0.908	1.013
Plate		1		2		3		4		5

Time	13:45	14:05	14:25	14:45
OD	1.083	1.136	1.192	1.253

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 5	Dilution 6	Dilution 7	5	6	7	
1	1	64	5	–	64	–	–	50 000 000
	2	35	3	–	35	–	–	
	3	51	4	–	51	–	–	
2	1	83	7	–	83	–	–	109 000 000
	2	118	9	–	118	–	–	
	3	126	16	–	126	–	–	
3	1	220	16	4	–	–	–	240 000 000
	2	250	24	4	–	24	–	
	3	252	24	6	–	24	–	
4	1	>300	40	–	–	40	–	456 666 000
	2	>300	41	–	–	41	–	
	3	>300	56	–	–	56	–	
5	1	>300	48	4	–	48	–	580 000 000
	2	>300	57	4	–	57	–	
	3	>300	69	6	–	69	–	

Table C2: Total viable count log phase day 2.

Time	10:40	11:00	11:40	12:00	12:20	12:40
OD	0.286	0.325	0.385	0.481	0.588	0.724
Plate	6	7				

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 5	Dilution 6	Dilution 7	5	6	7	
6	1	62	6	–	62	–	–	70 666 000
	2	71	6	–	71	–	–	
	3	79	10	–	79	–	–	
7	1	74	8	–	74	–	–	79 333 000
	2	77	8	–	77	–	–	
	3	87	14	–	87	–	–	

Table C3: Total viable count log phase day 3.

Time	10:05	10:30	10:50	11:10	11:40
OD	0.092	0.108	0.147	0.192	0.295
Plate	8	9	10	11	12

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 4	Dilution 5	Dilution 6	4	5	6	
8	1	127	13	–	127	–	–	13 566 000
	2	134	14	–	134	–	–	
	3	146	19	–	146	–	–	
9	1	149	8	–	149	–	–	17 166 000
	2	174	8	–	174	–	–	
	3	192	14	–	192	–	–	
10	1	–	20	1	–	20	–	21 500 000
	2	–	23	1	–	23	–	
	3	–	x	5	–	–	–	
11	1	–	42	4	–	42	–	50 000 000
	2	–	47	4	–	47	–	
	3	–	61	5	–	61	–	
12	1	–	80	6	–	80	–	77 000 000
	2	–	74	4	–	74	–	
	3	–	–	4	–	–	–	

Table C4: Total viable count log phase day 4.

Time	10:40	11:30	12:10	12:45
OD	0.312	0.425	0.652	0.869
Plate	13	14	15	16

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 4	Dilution 5	Dilution 6	4	5	6	
13	1	–	75	7	–	75	–	80 333 000
	2	–	81	7	–	81	–	
	3	–	85	8	–	85	–	
13	1	–	161	14	–	161	–	169 000 000
	2	–	168	15	–	168	–	
	3	–	178	16	–	178	–	
15	1	–	203	16	–	–	20	200 000 0000
	2	–	250	19	–	–	–	
	3	–	273	20	–	–	–	
16	1	–	–	43	–	–	43	420 000 000
	2	–	–	47	–	–	47	
	3	–	–	36	–	–	36	

Table C5: Total viable count log phase day 5.

Time	10:00	10:30	11:30	12:00	12:30	13:00
OD	0.055	0.067	0.121	0.201	0.309	0.414
Plate	17		18			19

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 4	Dilution 5	Dilution 6	4	5	6	
17	1	162	15	–	162	–	–	18 167 000
	2	186	17	–	186	–	–	
	3	197	17	–	197	–	–	
18	1	>300	x	–	–	–	–	No value
	2	>300	x	–	–	–	–	
	3	>300	x	–	–	–	–	
19	1	–	136	10	–	136	–	138 666 000
	2	–	139	14	–	139	–	
	3	–	141	14	–	141	–	

Table C6: Total viable count log phase day 6.

Time	12:40	13:20	13:50	14:10	14:25	14:55	15:20	15:40
OD	0.653	0.732	0.758	0.875	0.915	1.045	1.165	1.279
Plate	20		21	22	23			

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 5	Dilution 6	Dilution 7	5	6	7	
20	1	193	23	–	193	23	–	245 750 000
	2	221	26	–	–	26	–	
	3	221	30	–	–	30	–	
21	1	288	39	–	–	39	–	386 667 000
	2	317	40	–	–	40	–	
	3	325	47	–	–	47	–	
22	1	>300	42	–	–	42	–	465 000 0000
	2	>300	51	–	–	51	–	
	3	x	x	–	–	–	–	
23	1	>300	51	–	–	51	–	600 000 000
	2	>300	66	–	–	66	–	
	3	>300	63	–	–	63	–	

Table C7: Total viable count log phase day 7.

Time	9:30	9:55	11:00	11:25	11:55	12:25	12:45	13:05	13:25
OD	0.019	0.213	0.326	0.405	0.529	0.705	0.889	1.038	1.188
Plate	24	25	26	27	28	29			

Plate	**Repetition**	**Original count**			**Selected**			**CFU/mL**
		Dilution 4	Dilution 5	Dilution 6	4	5	6	
24	1	27	4	–	35	–	–	3 533 000
	2	38	5	–	37	–	–	
	3	41	6	–	51	–	–	
25	1	–	22	1	–	22	–	25 000 000
	2	–	26	3	–	26	–	
	3	–	27	3	–	27	–	
26	1	–	35	5	–	35	–	41 000 000
	2	–	37	6	–	37	–	
	3	–	51	7	–	51	–	
27	1	–	81	11	–	81	–	90 000 000
	2	–	88	13	–	88	–	
	3	–	101	14	–	101	–	
28	1	–	134	11	–	134	–	143 000 0000
	2	–	144	18	–	144	–	
	3	–	151	19	–	151	–	
29	1	–	201	20	–	–	20	250 000 000
	2	–	218	27	–	–	27	
	3	–	233	28	–	–	28	

Table C8: Total viable count log phase day 8.

Time	13:00	13:30	14:00	14:30	15:00	15:30	16:00	16:40
OD	0.811	1.081	1.215	1.331	1.448	1.508	1.534	1.487
Plate	30					31	32	33

Plate	**Repetition**	**Original count**		**Selected**		**CFU/mL**
		Dilution 6	Dilution 7	6	7	
30	1	46	4	46	–	520 000 000
	2	54	7	54	–	
	3	56	8	56	–	
31	1	72	7	72	–	753 333 0000
	2	74	10	74	–	
	3	80	15	80	–	
32	1	81	10	81	–	840 000 000
	2	87	11	87	–	
	3	x	4	–	–	
33	1	89	12	899	–	890 000 000
	2	89	12	89	–	
	3	x	11	–	–	

Table C9: Total viable count log phase day 9.

Time	13:50	14:15	14:40	15:15	15:50	16:00
OD	0.862	0.959	1.062	1.177	1.262	1.388
Plate	34		35	36	37	38

Plate	**Repetition**	**Original count**		**Selected**		**CFU/mL**
		Dilution 6	Dilution 7	6	7	
34	1	10	1	–	–	240 000 000
	2	20	3	20	–	
	3	28	4	28	–	
35	1	44	3	44	–	506 667 000
	2	50	4	50	–	
	3	58	7	58	–	
36	1	29	4	29	–	373 333 0000
	2	36	3	36	–	
	3	47	4	47	–	
37	1	39	2	39	–	406 667 000
	2	40	4	40	–	
	3	43	6	43	–	
38	1	36	2	36	–	393 333 000
	2	38	2	38	–	
	3	44	4	44	–	

- The result of plate 2 was discarded since other similar OD values showed that the concentration should be around 70×10^6 to 80×10^6 CFU/mL.
- The results achieved from day 8 were totally excluded from the whole analysis. The reason was based on the concentrations of *E. coli*. During log phases the concentration reached up to 8×10^8 and the concentrations after 24 hours were 1.2×10^9 which were much higher than any other experiments. A possible cause of the notable results might be some problems with pipette since the saline aliquots were prepared for all of them at once.

– The result of plate 35 was discarded because the concentration was anomalously high in comparison to others when they were implemented on the same batch of bacteria in the exponential phase within a day.

The following table demonstrates the TVC when the bacteria were maintained for 24 to 27 hours and moved to the stationary phase.

Table C10: Total viable count stationary phase.

Sample	**Time [h]**	**Rep.**	**Original count**			**Selected**			**CFU/mL**
			Dil. 5	Dil. 6	Dil. 7	5	6	7	
1	24	1	–	48	3	–	48	–	553 333 000
		2	–	51	8	–	51	–	
		3	–	67	12	–	67	–	
1	48	1	–	63	9	–	63	–	Contaminated
		2	–	66	10	–	66	–	
		3	–	67	12	–	67	–	
2	24	1	–	44	3	–	44	–	506 667 000
		2	–	50	4	–	50	–	
		3	–	58	7	–	58	–	
3	72	1	–	66	3	–	66	–	823 333 000
		2	–	82	5	–	82	–	
		3	–	99	7	–	99	–	
4	72	1	–	74	6	–	74	–	806 667 000
		2	–	80	9	–	80	–	
		3	–	88	10	–	88	–	
4	72	1	–	74	6	–	74	–	806 667 000
		2	–	80	9	–	80	–	
		3	–	88	10	–	88	–	
5	24	1	–	76	6	–	76	–	833 333 000
		2	–	84	9	–	84	–	
		3	–	90	x	–	90	–	
6	15	1	–	76	6	–	76	–	833 333 000
		2	–	84	9	–	84	–	
		3	–	90	x	–	90	–	
7	24	1	–	83	9	–	83	–	1 036 667 000
		2	–	101	14	–	101	–	
		3	–	127	16	–	127	–	

8	24	1	–	112	12	–	112	–	1 296 667 000
		2	–	129	15	–	129	–	
		3	–	148	x	–	148	–	
9	24	1	–	20	2	–	20	–	253 333 000
		2	–	26	2	–	26	–	
		3	–	30	3	–	30	–	
9	25	1	–	20	2	–	20	–	210 000 000
		2	–	21	2	–	21	–	
		3	–	22	2	–	22	–	
10	24	1	–	13	1	–	–	–	255 000 000
		2	–	22	2	–	22	–	
		3	–	29	2	–	29	–	
11	24	1	>300	49	–	–	49	–	510 000 000
		2	>300	53	–	–	53	–	
		3	>300	x	–	–	–	–	
11	25	1	>300	46	2	–	46	–	516 667 000
		2	>300	46	6	26	–	46	
		3	29	63	7	30	–	63	
11	26	1	–	51	–	–	51	–	586 667 000
		2	–	53	–	–	43	–	
		3	–	72	–	–	72	–	
12	24	1	>300	54	–	–	54	–	580 000 000
		2	>300	62	–	–	62	–	
		3	>300	x	–	–	–	–	
12	25	1	–	42	–	–	42	–	580 000 000
		2	–	53	–	–	53	–	
		3	–	79	–	–	79	–	
12	26	1	–	56	–	–	3	–	646 667 000
		2	–	60	–	–	6	–	
		3	–	78	–	–	10	–	

- Samples 7 and 8 demonstrated unusual high values of bacterial concentration. However, they were still able to confirm the stationary phase. Additionally, since they were measured on the same day with several other samples which also possessed much higher value of concentrations, the other values was discarded.
- Sample 9 and 10 were taken with OD number approximately 1.3. However, the OD number of the stationary phase and death phases varied around 1.5. Besides, the spectra behaved strongly different from the previous samples. Additionally, after two week, their TVC was noticeably low. Those mean there were some problems with the bacteria. Hence, the spectra of those samples were totally excluded from analysis process.

The following table demonstrates the TVC when *E. coli* was maintained for 7 days and 20 days, and were in the death phase.

Table C11: Total viable count death phase.

Sample	**Time [days]**	**Rep.**	**Original count**		**Selected**		**CFU/mL**
			Dilution 5	Dilution 6	5	6	
3	20	1	32	1	32	–	33 333 000
		2	33	1	33	–	
		3	35	2	35	–	
4	20	1	35	2	35	–	39 000 000
		2	40	3	40	–	
		3	42	5	42	–	
9	7	2	0	–	–	–	x
		2	0	–	–	–	
		3	0	–	–	–	
10	7	1	0	–	–	–	x
		2	0	–	–	–	
		3	0	–	–	–	
13	7	1	215	22	–	22	263 333 000
		2	240	24	–	24	
		3	248	33	–	33	
14	7	1	171	17	171	–	169 033 000
		2	218	20	–	–	
		3	230	29	–	–	